META-ANALYSIS

A Meta-Analysis of Mathematics Interventions: Examining the Impacts of Intervention Characteristics

Megan Rojo[1] · Jenna Gersib[2] · Sarah R. Powell[3] ·
Zhina Shen[4] ·
Sarah G. King[2] · Syeda Sharjina Akther[2] · Tessa L. Arsenault[2] ·
Samantha E. Bos[5] · Danielle O. Lariviere[2] · Xin Lin[6]

Abstract

The efficacy of mathematics interventions for students with mathematics difficulties is a critical concern in educational research, particularly in light of the increasing demand for effective instructional strategies. While previous meta-analyses have explored various aspects of mathematics interventions, our study uniquely focuses on a broad range of intervention characteristics and their impacts on students' mathematics proficiency. This meta-analysis reviewed 223 reports comprising 286 studies, encompassing data from approximately 56,477 participants. We analyzed several intervention characteristics as potential moderators, including (1) grade level, (2) group size, (3) content area, and (4) dosage. We also examined study level characteristics, including (1) research design, (2) fidelity, (3) year of study, (4) type of measure, and (5) study quality. Our findings revealed significant variability in intervention effectiveness based on targeted content areas, with fraction interventions demonstrating a pronounced impact on the effectiveness of interventions for students with MD. The study also underscored the challenges in educational research, marked by heterogeneity in type of measure and study designs. This comprehensive analysis contributes to the field by providing nuanced insights into the effectiveness of mathematics interventions, aiding educators and policymakers in developing strategies that effectively address the diverse needs of students with MD in various educational settings.

Keywords Mathematics difficulties · Intervention · Special education

Collaborative efforts since the 2008 National Mathematics Advisory Panel have aimed to improve mathematics education in the USA, leading to a consistent upward trend in mathematics performance among elementary and secondary students over the last decade (National Center for Education Statistics [NCES], 2022). However, a range of factors, including the COVID-19 pandemic, have posed significant

Extended author information available on the last page of the article

challenges toward this upward trajectory. According to the Nation's Report Card, there was a significant decline in the number of students demonstrating mathematics proficiency from the 2019 to the 2022 assessments, the largest drop between any two administrations of the National Assessment of Educational Progress (NCES, 2022). The average performance in 2022 was equivalent to that of 2005 results, erasing nearly 20 years of national progress. Performance decreased across all mathematics domains, including number properties and operations, measurement, geometry, data analysis, statistics and probability, and algebra. Moreover, 53% of fourth-grade students and 72% of eighth-grade students with disabilities scored below basic levels of proficiency. These findings highlight the persistent challenges in advancing mathematics education in the USA, notably for students with disabilities who have historically faced opportunity gaps compared to their typically achieving peers (Schulte & Stevens, 2015).

Despite these challenges, the unwavering dedication to bolstering mathematics education in the USA remains paramount. Otherwise, failing to establish proficiency in mathematics at an early age carries significant repercussions, particularly for students with mathematics difficulties (MD). Students with MD make up as much as 35% of the school population (Gersten et al., 2005a, 2005b). These students may have a school-identified Specific Learning Disability with Individualized Education Program goals in mathematics. However, many students with MD do not have a disability identification. Instead, they demonstrate mathematics performance that is below grade-level expectations and may be at risk for a learning disability. Like much of the research in the area of mathematics intervention (e.g., Fuchs et al., 2010; Jitendra et al., 2013), we use the umbrella term MD to encompass students who experience difficulty with mathematics with and without identification of a disability. Students with MD in high school exhibit heightened susceptibility to dropping out (Soland, 2013), hindering the pursuit of post-secondary education (Kim et al., 2015). These individuals may find themselves excluded from various career paths, particularly high paying opportunities in STEM fields (Wang et al., 2017). This underscores the need for targeted and effective interventions from K to 12, particularly considering the distinct challenges and diverse needs of this group.

The body of literature in special education offers robust insights into evidence-based instructional practices, which teachers can implement to enhance outcomes for students with MD. The What Works Clearinghouse, together with expert panels, has synthesized mathematics intervention research, resulting in the publication of practice guides offering instructional recommendations to improve student outcomes. The latest practice guide led by Fuchs et al. (2021) presented six concrete recommendations for teachers: (a) delivering systematic instruction, (b) using clear and concise mathematics language, (c) utilizing well-chosen mathematics representations to support student learning, (d) using number lines to build concepts and procedures, (e) providing targeted instruction on word problem-solving, and (f) incorporating timed activities to build mathematical fluency. These six recommendations serve as a guide for teachers in customizing their mathematics interventions to effectively meet their struggling students' needs.

Nonetheless, the landscape of learning is influenced by a multitude of factors beyond mere instructional strategies, which may significantly affect student

learning (Kroesbergen et al., 2003; Morgan & Sideridis, 2006). These external factors, or intervention characteristics (Table 1), include elements such as grade-level, instructional group size, mathematics content, and dosage. While these characteristics are crucial elements that can be manipulated by educators to optimize instructional intensity, there is a gap in understanding how these factors specifically impact student outcomes in mathematics. This highlights a significant need for research focused on understanding which intervention characteristics move the dial, particularly for students with MD who require instructional efficiency to close the achievement gap. This study aims to determine the significance of the impact of these intervention characteristics on student outcomes, a focus that has potential for influencing both teacher planning and decision-making as well as broader school-level policies.

Intervention Characteristics

Grade Level

Interventions targeting specific grade-levels may yield varying effect sizes due to the developmental nature of mathematical cognition (Dennis et al., 2016; Lein et al., 2020) and the foundational nature of certain mathematical concepts (Baroody & Lai, 2022). Early grade levels, for instance, focus on core numeracy skills which form the foundation for more complex mathematical operations taught in later years (Jordan et al., 2012). Intervening at these foundational stages is hypothesized to significantly alter the trajectory of students' mathematical capabilities (Gersten et al., 2005a, 2005b). Conversely, interventions in higher grade levels might target more complex mathematical concepts, and while critical, may build upon foundational concepts that students with MD are lacking. A meta-analysis conducted by Kroesbergen and Van Luit (2003) revealed that interventions for younger children had higher effect sizes than those for older children, which might be attributed to the malleable nature of early cognitive development. Yet, a more recent study conducted by Jitendra et al. (2021) found that, although nonsignificant, effect sizes were slightly higher for students in grades 6 through 8 and 9 through 12 versus grades K through 2 and grades 3 through 5. Schools and teachers can strategically utilize the results of this analysis to allocate resources more effectively, focusing on specific grade levels where interventions have been shown to yield greater gains, thus maximizing the impact of educational support for students with MD.

Group Size

Group size in educational interventions refers to the number of students involved in a learning session. The impact of group size on educational outcomes is a topic of considerable interest, particularly in special education and interventions for students with learning difficulties (Clarke et al., 2017; Doabler et al., 2019). Smaller groups are often theorized to offer more personalized attention and greater opportunity for student engagement, potentially leading to better outcomes for students with MD

Table 1 Operational definitions of covariates

Covariate	Operational definition
Grade level	We coded studies based on the reported grade level into three categories: (a) Early elementary: kindergarten–grade 2 (b) Upper elementary: grades 3–5 (c) Secondary: grades 6–12
Group size	We coded the size of intervention groups into four categories: (a) Individual interventions: groups of one student (b) Small group interventions: groups of two to five students (c) Large group interventions: groups of six or more students (d) Mixed/not reported: mixed grouping or studies who did not report group size
Content	We coded the content area in which the intervention primarily focused into eight categories based on the CCSS-M (Common Core, 2010) domains: (a) Algebra (b) Early numeracy (c) Fractions (d) Geometry (e) Measurement and data (f) Operations (g) Problem-solving (h) Mixed: interventions that focused on multiple content areas
Dosage	We coded dosage into four categories by multiplying the number of minutes per session by the total number of sessions students received the intervention. Based on our sample data, we included four categories: (a) 540 min or less (9 h) (b) 541–1020 min (9–17 h) (c) 1021–1350 min (17–22.5 h) (d) > 1350 min (22.5 h)
Year	We coded the year the study was published or submitted (as with dissertations). Based on our data, we split publication year into two groups: (a) Studies published between 2005 and 2012 (b) Studies published after 2012
Study Design	We coded the study design into three categories based on the experimental design described: (a) Quasi-experimental designs or RCT with high (> 30%) attrition (b) Randomized controlled trials (c) Single-case research designs
Measure Developer	We coded developer of outcome measures into two categories: (a) Researcher-developed: measures created by the author team (b) Standardized: commercially available or norm-referenced measures
Fidelity	We coded for the presence of reported implementation fidelity using four categories based on the CEC Quality Indicators (Cook et al., 2015). We looked for the following components: fidelity information pertaining to dosage, adherence, and frequency of fidelity reporting (i.e., if fidelity was recorded regularly throughout the intervention for each intervention agent). Then, we coded how many forms of fidelity were reported. (a) None reported (b) One reported (c) Two reported (d) Three reported

Table 1 (continued)

Covariate	Operational definition
Quality	We calculated the quality of each study using the CEC quality indicators (Cook et al., 2015) and divided the number obtained by the number possible. We grouped quality scores into three categories, based on our sample (a) Studies with a quality score less than 75% (b) Studies with a quality score between 75–90% (c) Studies with a quality score greater than 90%

(Fuchs et al., 2017). However, the literature presents mixed findings on the influence of group size. A meta-analysis conducted by Jitendra et al. (2018) suggested that small group instruction can be particularly effective for secondary students with learning disabilities. Conversely, a meta-analysis by Stevens et al. (2018) indicated that while small group learning can be beneficial for upper elementary and secondary students, it does not always surpass the effectiveness of whole-class instruction. For students with MD, the ideal group size may vary depending on the specific needs of the students and the nature of the intervention being employed. Thus, understanding the role of group size in mathematics interventions is crucial for tailoring instructional approaches to maximize student learning and achievement.

Content Area

Mathematics proficiency involves a comprehensive understanding and application of concepts and skills across several domains. These domains include (a) whole number properties and operations, (b) fractions, (c) measurement and data, (d) geometry, and (e) algebra (National Governors Association Center for Best Practices & Council of Chief State School Officers [NGAC & CCSSO], 2010). Given the intricate nature of these domains, mastery in each is crucial for comprehensive mathematical understanding. In addition to the domains listed above, we included early numeracy and problem-solving as separate domains because they have garnered significant attention in the literature for their role in building students' mathematics proficiency (Fuchs et al., 2021; Gersten et al., 2005a, 2005b; Purpura et al., 2011) and, thus, have been the foci of many intervention studies (e.g., Morin et al., 2017; Nozari et al., 2021; Shanley et al., 2017). While early numeracy and problem-solving interventions may include one or more of the mathematics domains listed above, their central focus is on building initial whole number concepts and applying mathematics concepts in word problems, respectively.

Existing meta-analyses, when evaluating the impacts of interventions across content areas, present divergent findings. For instance, Stevens et al. (2018) pinpointed fractions interventions as being more effective than other content areas at improving student outcomes. Contrarily, studies from both Jitendra et al. (2018) and Williams et al. (2022) suggested that all mathematics content areas resulted in comparably effective outcomes. Consequently, this study seeks to elucidate the disparities in outcomes based on mathematics content area to discern if certain domains boast more developed and refined interventions relative to others. Understanding the nuances

in mathematics content areas, as revealed by this analysis, may be academically valuable for developing targeted, effective curriculum and instructional strategies in schools. The subsequent section will explore the significance of each domain in achieving proficiency in mathematics, and address the specific obstacles encountered by students with MD in mastering these areas.

Early Numeracy Before formal schooling in kindergarten, students begin learning early numeracy skills. Early numeracy encompasses various skills, such as counting, recognizing numbers, understanding numbers, comparison of numbers, addition concepts, and subtraction concepts, among others (Jordan et al., 2012; Toll & Van Luit, 2012). The Common Core State Standards (CCSS) emphasize early numeracy knowledge in kindergarten, which covers counting, comparison, place value, addition, and subtraction (NGAC& CCSSO, 2010). Early numeracy is the foundation for later mathematics learning (Clements & Sarama, 2013). For example, preschool and kindergarten early numeracy scores predicted mathematics scores in grade 1 (Missall et al., 2012), grade 3 (Jordan et al., 2009), grade 5 (Hannula-Sormunen et al., 2015), and grade 8 (Geary et al., 2013).

Differences in mathematics from student to student emerge quite early. Before kindergarten, researchers have demonstrated that some students perform better on mathematics tasks than other students (Chu et al., 2013; Mazzocco et al., 2011; Watts et al., 2014). At kindergarten, researchers have identified students with MD (Clarke et al., 2011; Dyson et al., 2011) to provide early intervention to these students. In studies, students with MD have demonstrated difficulty with counting, which is foundational to many other skills in mathematics (Aunio et al., 2015; Koponen et al., 2013), as well as comparison and subitizing (Raddatz et al., 2017; Rousselle & Noël, 2007). Furthermore, researchers have noted that students with MD in kindergarten continue to show below-average mathematics performance in later grades (Jordan et al., 2009; Stock et al., 2010).

Whole Numbers As students develop an understanding of early numeracy, they learn about whole numbers and the operations of addition, subtraction, multiplication, and division. Emphasizing whole numbers, addition, and subtraction is prevalent in mathematics standards from kindergarten to grade 2, with multiplication and division entering the standards in grade 3 (NCAC & CCSSO, 2010). As whole number understanding is essential for success with fractions (Hansen et al., 2015), the development of proficiency with whole numbers and operations in the elementary grades may set students up for success with mathematics in later grades.

Similar to early numeracy, students with MD also experience difficulty with whole numbers and operations. As examples, students with MD have demonstrated lower performance on tasks related to facts (Andersson, 2008; Cirino et al., 2015; Powell et al., 2009) and computation (Chong & Siegel, 2008; Lambert & Moeller, 2019; Mancl et al., 2012; Nelson & Powell, 2018; Vukovic & Siegel, 2010). Furthermore, students with MD showed difficulty with word-problem solving, which often relies on whole number and operation knowledge (Cowan & Powell, 2014; Reikerås, 2009). Because whole numbers and operations are predictors of later mathematics

performance (Bailey et al., 2014; Siegler et al., 2012; Vukovic et al., 2014), this content may be foundational to mathematics intervention efforts in the elementary grades.

Fractions Fraction knowledge becomes a focus in grade 3 (NCAC & CCSSO, 2010) with early fraction instruction focused on understanding fractions then comparing and ordering fractions. By the end of grade 5, students are expected to add, subtract, multiply, and divide fractions. Fractions are often considered essential for later success with algebra (Booth & Newton, 2012). In fact, Siegler et al. (2012) identified fraction knowledge, along with division, as a unique predictor of high school mathematics scores. Cirino et al. (2019) noted a similar pattern in which fractions scores in grade 6 predicted algebra performance in grade 9.

Students with MD also demonstrate performance below peers on fraction tasks (Namkung & Fuchs, 2016; Tian & Siegler, 2017), which is not surprising given that knowledge of whole numbers and operations predicts performance on fraction measures (Hansen et al., 2015; Vukovic et al., 2014). Of note, students with MD continue to experience difficulty with fractions across grade levels, with limited growth on fraction knowledge from grade 3 to 6 (Jordan et al., 2017).

Geometry Geometry is an essential component of the K-12 mathematics curriculum (NCAC & CCSSO, 2010; Mistretta, 2000). According to the common core state standards of mathematics, students need to learn geometric shapes practice spatial reasoning, apply transformations, use visualization, and utilize geometric modeling problem-solving (National Council of Teachers of Mathematics, 2000). To improve overall mathematics performance, students need to master geometry in addition to other mathematics contents (e.g., whole numbers, and fractions). As students' progress to high school, the level of complexity in geometry increases (Zhang, 2021), and a solid foundation in geometry is essential for students planning to pursue college studies in science, technology, and engineering fields (Chen et al., 2021).

Students with MD can have more difficulties in geometry when compared with other mathematics content. The trends in International Mathematics and Science Study (TIMSS), which assesses proficiency in algebra, number, data and chance, and geometry reported that US grade 8 students perform significantly lower in geometry compared to other content areas (Gonzales et al., 2004; Provasnik et al., 2012). Additionally, the most recent report from the NAEP in 2022 indicated that average scores in geometry have declined in 42 out of 50 states since 2019 (NCES, 2022).

Measurements and Data Analysis Measurement and data analysis are integral parts of mathematics, and students in every grade level are expected to improve proficiency in this area of mathematics (NGAC & CCSO, 2010). Research indicates that students who have a robust understanding of early measurement and data analysis tend to perform stronger in later mathematics (Frye et al., 2013; Sarama & Clements, 2009). A better understanding of these areas facilitates students to apply and strengthen their whole number understanding and complex statistical investigation,

such as posing important research questions that can be answered through data analysis (Doabler et al., 2019; Frye et al., 2013). Furthermore, foundational knowledge of measurement and data analysis provides the opportunity for students to develop advanced mathematics skills, which are needed for future STEM fields (Clements et al., 2020; Doabler et al., 2022).

Despite its importance, students with MD exhibit difficulty in reaching proficient levels in measurement and data analysis. For example, according to the NAEP data (2022), students' scores in measurements and data analysis decreased by 4–5 points in grade 4, and 9–10 points in grade 8 compared with the 2019 data. A potential source of difficulty may be in interpreting various forms of data representations, such as bar graphs, picture graphs, and line plots. In elementary mathematics education, students are frequently required to engage in measurement and data interpretation while solving word problems. Research indicates that students with MD often struggle with understanding and interpreting these data, which poses a significant challenge to effectively solving word problems that incorporate measurement and data analysis (Doabler et al., 2019).

Problem-Solving Problem-solving in mathematics education refers to the ability to use mathematical concepts and procedures to solve real-world problems (NCAC & CCSSO, 2010). It is an essential skill, as it transcends rote learning and involves critical thinking and application of mathematical knowledge. For students with MD, problem-solving can be particularly challenging due to deficits in foundational math skills, working memory, and reasoning abilities (Fuchs et al., 2009). Research indicates that problem-solving interventions can significantly enhance mathematical outcomes for these students. A meta-analysis by Lein et al. (2020) found that interventions focusing on problem-solving strategies led to improved performance in students with MD. However, the complexity of problem-solving tasks requires tailored instructional methods that align with students' specific needs.

Dosage

Dosage in the context of educational interventions refers to the intensity and duration of the instructional intervention (Powell & Fuchs, 2015). Higher dosages, characterized by longer duration or greater frequency of instruction, are hypothesized to lead to better educational outcomes (Fuchs et al., 2017). For students with MD, sufficient dosage is crucial for reinforcing learning and ensuring mastery of mathematical concepts. Research by Wanzek et al. (2016) suggested that interventions with extended duration and higher intensity are more effective in improving academic skills in students with learning difficulties. However, a meta-analysis of mathematics word problem-solving interventions conducted by Lein et al. (2020) determined that instructional minutes had no impact on overall effect sizes. The findings on dosage can inform schools and teachers about optimizing intervention duration and intensity, offering guidance on balancing instructional time for maximum effectiveness in supporting students with MD.

Study Characteristics

In addition to the intervention characteristics that schools and teachers can consider when making instructional decisions and policies, it is essential to recognize the possible influence that study characteristics may have on mathematics outcomes (Polanin et al., 2019). These study characteristics are inherent to the research process and methodology and may not be directly hypothesized to impact mathematics outcomes in school settings; therefore, they will be held constant within our model to facilitate a more comprehensive analysis. Table 1 lists the study characteristics hypothesized to influence outcomes, including (a) the year the study was conducted, (b) the research design employed, (c) the type of outcome measure used, (e) the fidelity of implementation in the research studies, and (f) the study quality.

Research Design

The research design employed in study may influence the magnitude of the effect sizes (Zeneli et al., 2016). Specifically, randomized controlled trials (RCT), quasi-experimental designs (QED), and single-case research designs (SCRD) may have differences related to their designs that affect internal validity or the magnitude of the outcome (Kraft, 2020; Lipsey & Wilson, 1993). SCRD are frequently utilized in special education, focusing on the effects of an intervention on an individual and may be particularly valuable in identifying functional relations (Maggin et al., 2018). On the other hand, RCTs are often heralded as the "gold standard" in experimental research due to their rigorous random assignment, which minimizes biases and confounding variables (Gersten et al., 2000). RCTs provide a high degree of internal validity but might report more conservative effect sizes due to their rigorous random assignment (Cheung & Slavin, 2016). In QED, internal validity is compromised due to a lack of random assignment (Gersten et al., 2005a, 2005b). Such designs may inflate effect sizes based on the comparability of the groups (Cheung & Slavin, 2016). Prior meta-analyses examining research design as a potential moderator have yielded mixed results. Some studies have found SCRD to report larger effect sizes compared to RCTs (e.g., Dessemontet et al., 2019; Losinski et al., 2014), while others found no differences between research designs (e.g., Jitendra et al., 2017; Myers et al., 2023; Rojo et al., 2023).

Type of Measure

The type of measure used to assess student performance is an important consideration when interpreting effect sizes in education (Lipsey et al., 2012). Researcher-developed measures are often more closely aligned to the research question being investigated, but they may have lower reliability and validity than standardized measures (Wolf & Harbatkin, 2022). Standardized measures, on the other hand, have been normed on a large sample of participants and have established psychometric properties (Rumrill et al., 2020), but they may be more distal to the intervention.

Researchers may have valid reasons for choosing to use either researcher-developed or standardized measures, or a combination of both, depending on the specific research questions being addressed (Toste et al., 2023). However, it is important to control for the influence that the type of measure may have on effect sizes when comparing across studies because researcher-developed measures are more likely to yield larger effect sizes than standardized measures, even when the interventions being compared are equally effective (Cheung & Slavin, 2016).

Fidelity

Implementation fidelity in educational research refers to the degree to which interventions are delivered as intended (Hill & Erickson, 2019). High fidelity is critical to ensure the reliability of intervention outcomes. Despite irregular reporting of fidelity in educational research (Swanson et al., 2013), studies such as O'Donnell (2008) found that fidelity of implementation was positively associated with student outcomes. In contrast, a systematic review by Rojas-Andrade and Bahamondes (2019) revealed that components of implementation fidelity were linked to positive outcomes only 40% of the time. Therefore, it is essential to explore the influence that implementation fidelity may have on the effectiveness of mathematics interventions.

Year Band

The year band of the study can be a significant moderator in educational research, reflecting changes in educational policies, practices, and technologies over time (Li et al., 2020; Odden et al., 2021). Studies conducted in different time periods may capture different educational contexts and challenges. For instance, interventions post-2012 may incorporate more tablet-mediated mathematics instruction and align with newer educational standards compared to those before 2012 (Liu et al., 2023). Research by Raudenbush (2008) highlights the importance of considering temporal factors in educational research, as they can provide insights into the evolution and effectiveness of instructional strategies.

Study Quality

Study quality in educational research refers to the methodological rigor and reliability of the studies included in a meta-analysis (Talbott et al., 2018). High-quality studies are characterized by robust research designs, valid and reliable measures, and clear reporting of results (Gersten et al., 2005a, 2005b). Assessment of study quality is the systematic evaluation of how well a study adheres to the highest possible quality standards (Higgins & Green, 2008). Assessing study quality is vital for drawing accurate conclusions from a meta-analysis (Cook et al., 2015; Protogerou & Hagger, 2020). As Slavin and Smith (2009) indicated, studies with strong research designs yield more reliable effect sizes. Quality indicators in special education research are designed to ensure that evidence-based practices are supported by a rigid body of literature and are used to assess study quality (Cook et al., 2014).

Meta-analyses of mathematics interventions have included study quality as a moderator (e.g., Dennis et al., 2016; Jitendra et al., 2021; Lei et al., 2020; Myers et al., 2021; Stevens et al., 2018), but findings have been mixed. Dennis et al. (2016) and Lei et al. (2020) reported a statistically significant moderation effect of study quality on the efficacy of mathematics interventions, and they found comparable results indicating studies with higher quality were found to have lower effect size estimates. Conversely, other researchers (Jitendra et al., 2021; Myers et al., 2021; Stevens et al., 2018) found that study quality did not significantly moderate the effect of mathematics interventions. Our meta-analyses will provide nuanced insights and add to the existing literature on the role of study quality in educational research.

Purpose and Research Questions

We conducted this study to address some of the limitations from previous meta-analysis and to provide a comprehensive understanding of mathematics interventions for students with MD. In doing so, we adopted a broader inclusion criterion, encompassing a wide range of studies. This approach allowed us to capture a more comprehensive dataset, enhancing our ability to assess the impact of various moderators more effectively than previous analyses. This is of particular importance because schools may need more supplemental mathematics support than they have in the last few decades, primarily because of COVID-19 school disruptions. By including studies that cover a wide spectrum of intervention characteristics and settings, our meta-analysis is uniquely positioned to explore optimal conditions for developing mathematics proficiency. This will enable schools to implement interventions more effectively in a post-pandemic era, addressing the needs of students with MD in a manner that previous studies may not have fully captured. Our research questions included the following:

1. What is the estimated average effect of mathematical interventions on mathematical performance for K-12 students with MD?
2. To what extent does the efficacy of mathematical interventions vary based on intervention characteristics (i.e., grade level, group size, content area, dosage) or study characteristics (i.e., research design, measure type, fidelity, publication year, and study quality)?

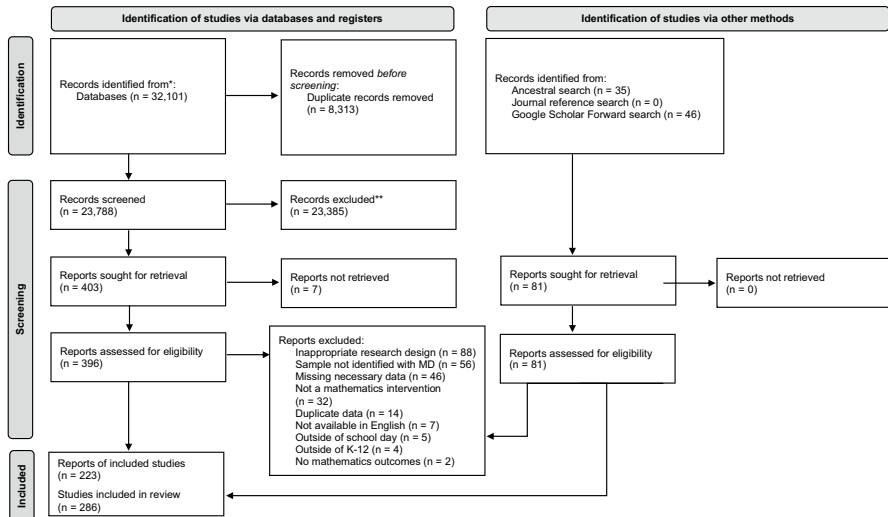## Method

Search Procedures

We conducted a search of peer-reviewed journals and gray literature (i.e., dissertations and theses) in these electronic databases: Academic Search Complete, Education Source, Educational Resources Information Clearinghouse (ERIC), PsycINFO, and ProQuest Dissertations and Theses Global. The search included reports published between January 2005 and December 2020. We selected 2005 as the start date

of the search because it follows the publication of the quality indicators for group design and SCRD (Gersten et al., 2005a, 2005b; Horner et al., 2005).

We used variants of the following search terms in our search. The first set of search terms were focused on capturing studies of mathematics interventions and included mathematics, arithmetic, geometry, algebra, calculus, numeracy, word problems, and story problems. The second set of terms captured our population (students with MD). The terms we used included learning disabilities, learning disorders, learning difficulties, at risk, special education, learning problem, mathematics difficulty, mathematics disabilities, and dyscalculia. Finally, the last set of terms captured interventions and included program, intervention, tutor, instruction, tier 2, tier 3, curriculum, lesson, treatment, differentiated learning, educational strategy, teaching method, and schema. The initial screening process yielded 32,101 abstracts. After removing duplicates, we screened the abstracts of 23,788 records. Through this process, we identified 403 reports that met our initial inclusion criteria, but seven were not retrieved. We conducted an ancestral search from relevant reviews (i.e., Bouck et al., 2018; Chodura et al., 2015; Codding et al., 2011; Dennis et al., 2016; Ennis & Losinski, 2019; Gersten et al., 2009a, 2009b; Jitendra et al., 2017; Lee et al., 2020; Lei et al., 2020; Powell et al., 2020; Stevens et al., 2018; Zhang & Xin, 2012), which resulted in an additional 35 reports. We completed a 15-year hand search of the following special education journals: *Exceptional Children*, *Journal of Learning Disabilities*, *Journal of Special Education*, *Remedial and Special Education*, *Learning Disabilities Quarterly*, and *Learning Disabilities Research and Practice* and found no additional reports. After reviewing the full text of these 431 reports, we identified 177 reports that met our full inclusion criteria. We conducted forward and backward searches (i.e., examined reference lists) on all 177 reports and identified an additional 46 reports that met initial inclusion criteria, totaling 223 reports. Some reports ($k=54$) contained data for two or three experimental treatments, yielding a total of 286 experimental studies for this meta-analysis. Figure 1 displays a Preferred Reporting Items for Systematic Review and Meta-Analyses diagram describing the search process (Page et al., 2020).

Inclusion Criteria

We included studies that met the following inclusion criteria: (1) The study employed an experimental design, including a randomized group, quasi-experimental, or SCRD design. (2) The study included students in kindergarten through grade 12 who experienced learning difficulties with mathematics (i.e., MD). We identified MD based on author descriptions of below-average mathematics performance (i.e., a diagnosis of a mathematical learning disability, at-risk status for mathematics learning disability). Studies with combined samples of participants (i.e., students with and without MD or students with MD and other disabilities) were included if data were disaggregated for students identified with MD or at least 50% of the sample included students with MD. (3) The authors implemented a mathematics intervention and measured students' post-intervention mathematics performance. We defined *intervention* as instruction provided related to mathematics that occurred

Identification of studies via databases and registers

Records identified from*:
Databases (n = 32,101)

Records removed *before screening*:
Duplicate records removed (n = 8,313)

Records screened
(n = 23,788)

Records excluded**
(n = 23,385)

Reports sought for retrieval
(n = 403)

Reports not retrieved
(n = 7)

Reports assessed for eligibility
(n = 396)

Reports excluded:
Inappropriate research design (n = 88)
Sample not identified with MD (n = 56)
Missing necessary data (n = 46)
Not a mathematics intervention (n = 32)
Duplicate data (n = 14)
Not available in English (n = 7)
Outside of school day (n = 5)
Outside of K-12 (n = 4)
No mathematics outcomes (n = 2)

Identification of studies via other methods

Records identified from:
Ancestral search (n = 35)
Journal reference search (n = 0)
Google Scholar Forward search (n = 46)

Reports sought for retrieval
(n = 81)

Reports not retrieved
(n = 0)

Reports assessed for eligibility
(n = 81)

Reports of included studies
(n = 223)

Studies included in review
(n = 286)

Identification | *Screening* | *Included*

Fig. 1 PRISMA 2020 Flow Diagram. *From:* Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. https://doi.org/10.1136/bmj.n71. For more information, visit: http://www.prisma-statement.org/

during the school day. We did not include interventions that took place as summer courses or in remedial programs outside of school. (4) The study was published in English between January 2005 and December 2020, to encapsulate mathematical intervention data collected prior to the COVID-19 pandemic.

Coding Procedure and Reliability

We developed a coding manual and a code sheet to extract relevant information from the reports that met inclusion criteria. Two of the members of the author team trained the remaining authors on the code sheet. To ensure high inter-rater reliability (IRR), all authors independently coded three reports (Bryant et al., 2021; Flores et al., 2014; Tournaki et al., 2008), met to discuss discrepancies and areas that required clarification, and then refined the code sheet to enhance coding reliability. This process was repeated as necessary for each individual coder. Once all members had an IRR score of at least 90%, each coder began coding reports independently.

For each study, we extracted the following information: (a) research design: type of design (i.e., RCT, QED, SCRD), random assignment (i.e., class level, individual level); (b) participants (i.e., identification of MD, grade level, age, number of participants with MD, disability label or classification, status as an English learner, gender, race, and ethnicity); (c) implementors (i.e., number of implementors, training, and qualifications); (d) intervention (i.e., sample size, number of sessions and duration, mathematics content, language of the intervention, instructional strategies, components, group size); (e) counterfactual (i.e., description of the counterfactual;

designation as active or passive control); (f) implementation fidelity (i.e., who implemented intervention, inclusion of theory of change, treatment differentiation, whether fidelity data included in analysis); (g) measures (i.e., dependent variables included, maintenance data, generalization data); (h) effect size information based on the What Works Clearinghouse 2020 guidelines (i.e., means and standard deviations or *f*-tests, sample sizes); and (i) quality indicators (i.e., study inclusion of essential methodological elements of special education, as clarified by Cook et al., 2015). For the quality indicators, we calculated a total percentage for each study as the number of included quality indicators divided by the total number of indicators possible (22 for SCRD, 24 for group designs) based on the CEC Quality Indicators (Cook et al., 2015). We present the coding criteria and operational definitions of each intervention and methodological design characteristic we included in the meta-regression in Table 1. To determine the reliability of our coding, authors independently checked 18% of the other authors' codes. We calculated IRR as the agreements divided by the sum of the agreements and disagreements, multiplied by 100. IRR was 89.1%, and we resolved all discrepancies to ensure the final coding sheet included 100% accuracy of coding.

Effect Size Calculation

The standardized mean difference (SMD) was calculated for each experiment's outcome(s) when sufficient information was reported, specifically focusing on the measurement of mathematical performance. The calculation of the SMD involved subtracting the mean of the control group from the mean of the treatment group and dividing the result by the pooled standard deviation (Cooper et al., 2019). To address the issue of small samples sizes within certain studies, a correction known as Hedges' *g* was applied to each SMD (Hedges & Olkin, 1985). An *f*-test was employed to calculate Hedges' *g* when means and standard deviations were not reported. Furthermore, a difference-in-difference adjustment was used for RCTs with high attrition and QED to account for pretest means (What Works Clearinghouse, 2020). For cluster RCTs, Cohen's *d* was calculated using the computer program Computing Effect Sizes for Cluster Randomized Studies (See Cooper et al., 2019, p. 241) and then converted to Hedges' *g*.

For SCRDs, ($k = 86$), we calculated the between-case SMD (BC-SMD), which is interpreted comparatively to Hedges' *g* (Shadish et al., 2014). The BC-SMD was calculated for the following SCRD designs: (a) multiple-baseline across participants, (b) multiple-probe across participants, (c) multiple-baseline across settings or behaviors (with a minimum of three replications), and (d) reversal designs. Studies that did not include one of these designs were not included in the meta-analysis ($k = 10$), as no comparative effect size could be calculated. The web-based Shiny app developed by Pustejovsky et al. (2021) was employed for the calculation of BC-SMD.

Meta-Analytic Models

To calculate the estimated average effect size across all studies, we ran an intercept-only model without the inclusion of covariates. This model was implemented using the *metafor* package in R (Viechtbauer, 2010). To investigate sources of heterogeneity within mathematic interventions for students with MD, we ran a series of meta-regression models. We calculated mixed-effect meta-regression models to observe variation in effect sizes across a series of covariates. The mixed effect model included fixed effect variability from each covariate and random effects of residual heterogeneity and within-study variance that resulted from outcome measures with possible dependencies (Borenstein et al., 2009). To account for dependencies within studies, we used robust variance estimation, which incorporates Sattherthwaite degrees of freedom and standard errors for regression coefficients to correct for small samples (Tipton & Pustejovsky, 2015). To calculate the effects, we used the *metafor* package in R, which accounts for both the hierarchical (effects within studies) and multivariate (multiple effects on the same population of students) nature of the data (Viechtbauer, 2010). We also used the *clubSandwich* package to run the robust variance estimation in R (Pustejovsky, 2018). Based on the guidelines by Pustejovsky and Tipton (2022), we assumed a correlation of 0.80 for effect sizes with multiple outcomes within the same sample. To determine the robustness of the model, we also tested correlations at 0.6 and 0.9. The data reported similar outcomes across each, indicating our mode was robust with each assumed correlation tested. Using robust variance estimation, type I error is inflated when Satterthwaite degrees of freedom are less than four. Thus, any covariates with degrees of freedom less than four are reported as unreliable (Fisher & Tipton, 2015).

For the meta-regression, we coded the following variables categorically and, thus, included them as factors within our meta-regression: grade level, group size, content area, dosage, year of publication, study design, measure developer, reported implementation fidelity, and study quality.

Publication Bias

We assessed our samples with and without outliers for publication bias using the Egger's test with robust variance estimation, as suggested by Pustejovsky and Rodgers (2019). Results of the model inclusive of all studies were significant ($p < 0.001$), signifying an asymmetric distribution of effect sizes. The results for the model excluding significant outliers were nonsignificant ($p > 0.05$), signifying (a) that the distribution of effect sizes was symmetric in the sample that excluded outliers, and (b) that the outliers significantly affected our findings. Based on these results and our inclusion of gray literature, we conclude that there is no evidence of publication bias in our model excluding outliers. Given the significant differences between the models with and without outliers, we present the findings from both models in our results.

Results

Our goal was to examine the overall efficacy of mathematical interventions for students with MD. We found 223 reports, 286 studies, and 825 total effect sizes examining mathematical outcomes for students with MD. These data reflect the mathematical progress of nearly 56,477 participants ($M = 71.6$ per study) receiving over 2800 h of mathematics instruction. Approximately 49% of the sample population identified as female, 25.3% as White, 19.4% as Latino/a, 15.6% as Black, 0.03% as multiracial or other, 0.02% were Asian, and less than 0.01% were Native American or Pacific Islander. Approximately 40% of studies included students with MD that did not have an identified disability. We also calculated the methodological quality of each study, resulting in an average of 81.6% quality indicators met ($SD = 13.1\%$), indicating a need for improvement in the overall quality of mathematics intervention studies.

Overall Efficacy

Our first research question aimed to determine the impact of mathematics interventions for students with MD. The average estimate effect of our sample was $g = 1.29$ ($SE = 0.14$, $p < 0.001$). As part of our analyses, we completed visual and statistical analyses to determine whether any effects skewed our results. We examined scatterplots and measures of skewness and kurtosis, finding significant skew in our data. Next, we completed the Tukey's outlier detection test, which considers outliers as data points that exceed the 75th percentile by a factor of 1.5 times the interquartile range, which, for our data, was effect sizes that exceeded -1.17 and 2.33 (Doane & Seward, 2011). Thus, we excluded 90 effect sizes from 37 studies and reran the analyses. The estimated average effect of mathematics interventions exclusive of outliers was $g = 0.56$ ($SE = 0.04$, $p < 0.001$). These results indicate that, while still large, there is a significant reduction in the overall estimated effect and a sizable reduction in the standard error across the 198 included studies, thus demonstrating the outliers significantly influenced our overall data set. Therefore, we report the remaining results with and without outliers in Table 2.

Moderators of Treatment Efficacy

Next, we examined potential moderators of treatment efficacy to determine the conditions in which mathematics interventions have been most beneficial. In the full model inclusive of outliers, our meta-regression did not reveal any significant moderators. On the model excluding outcomes, four moderators significantly accounted for the heterogeneity in our sample: problem-solving interventions ($\beta = -0.33$, $SE = 0.12$, $p < 0.01$), interventions involving operations ($\beta = -0.34$, $SE = 0.14$, $p < 0.05$), standardized measures ($\beta = -0.13$, $SE = 0.05$, $p < 0.05$), and SCRDs ($\beta = 0.45$, $SE = 0.11$, $p < 0.001$). Each covariate is discussed.

Table 2 Multiple meta-regression analyses with and without outliers

Moderator	Full model					Model excluding outliers				
	β	SE	df	p	95% CI	β	SE	df	p	95% CI
Intercept	0.44	1.07	47.96	.69	(−1.41, 2.28)	0.91	0.12	33.40	<0.001***	(0.65, 1.18)
Grade-level (vs. 3–5)										
K-2	−1.67	1.10	25.57	.14	(−3.65, 0.32)	−0.05	0.19	22.07	0.81	(−0.32, 0.23)
6–12	0.75	0.89	76.15	.40	(−0.70, 2.21)	−0.15	0.10	49.04	0.13	(−0.38, 0.07)
Group size (vs. small group)										
One	−0.05	0.57	38.21	0.93	(−1.07, 1.17)	0.00	0.10	51.13	0.93	(−0.18, 0.19)
Whole group	−2.27	1.40	47.20	0.11	(−4.03, −0.52)	−0.10	0.11	33.31	0.37	(−0.35, 0.15)
Mixed/NA	2.03	2.34	8.35	0.41	(−1.25, 5.31)	0.25	0.28	6.49	0.41	(−0.23, 0.73)
Content area (vs. fractions)										
Algebra	1.18	1.45	29.41	.42	(−1.52, 3.89)	−0.27	0.19	16.80	0.18	(−0.72, 0.18)
Early Num	1.03	1.42	53.70	.47	(−1.63, 3.69)	−0.43	0.22	51.07	0.06†	(−0.80, −0.05)
Geometry	3.02	1.94	2.64[1]	.23	(−1.90, 7.93)					
Measurement	8.34	6.44	3.06[1]	.28	(3.51, 13.17)	−0.37	0.26	1.61[1]	0.33	(−1.13, 0.40)
Operations	0.99	1.09	69.69	.37	(−0.83, 2.80)	−0.34	0.14	50.93	0.02*	(−0.62, −0.07)
Prob-solving	−0.38	1.00	70.34	.70	(−2.11, 1.34)	−0.33	0.12	53.62	0.007**	(−0.59, −0.07)
Mixed	0.60	1.70	17.99	.72	(−2.36, 3.55)	−0.15	0.15	16.13	0.33	(−0.56, 0.25)
Dosage (vs. >1350 min)										
<540 min	0.50	0.90	69.49	.58	(−0.90, 1.90)	−0.18	0.11	56.84	0.09†	(−0.40, 0.03)
541–1020 min	0.73	1.03	39.19	.48	(−0.58, 2.03)	−0.11	0.08	61.35	0.19	(−0.30, 0.08)
1021–1350 min	−2.03	1.33	10.01	.16	(−2.62, −1.45)	−0.16	0.11	4.39	0.11	(−0.33, 0.00)
Year (vs. <2013)										
2013–2020	0.57	0.68	75.24	.40	(−1.84, 0.70)	0.10	0.10	53.66	0.31	(−0.09, 0.29)

Table 2 (continued)

Moderator	Full model					Model excluding outliers				
	β	SE	df	p	95% CI	β	SE	df	p	95% CI
Study design (vs. RCT)										
QED	0.24	1.30	22.12	.85	(−2.07, 2.56)	0.06	0.22	21.16	0.77	(−0.29, 0.28)
SCRD	0.51	0.96	80.96	.60	(−0.85, 1.87)	0.45	0.11	58.89	<0.001***	(0.22, 0.67)
Type of measure (vs. researcher developed)										
Standardized	−0.16	0.53	4.43	.78	(−0.47, 0.15)	−0.13	0.05	27.79	0.02*	(−0.22, −0.04)
Implementation fidelity (vs. dosage, adherence, and frequency reported)										
None reported	0.42	0.30	1.20[l]	.37	(0.16, 0.67)	0.22	0.13	7.00	0.15	(0.07, 0.37)
One reported	6.95	4.37	20.80	.13	(6.29, 7.61)	−0.04	0.10	37.70	0.67	(−0.22, 0.13)
Two reported	−1.63	1.40	24.26	.25	(−2.19, −1.07)	0.05	0.10	29.24	0.58	(−0.09, 0.19)
Quality (vs. > 90%)										
<75%	−1.33	1.14	85.99	.25	(−2.86, 0.19)	−0.09	0.13	63.22	0.46	(−0.33, 0.14)
75–90%	0.39	0.64	121.24	.55	(−0.88, 1.66)	−0.03	0.09	77.61	0.73	(−0.22, 0.16)

RCT randomized controlled trial; *QED* quasiexperimental design; *SCRD* single-case research design; x^1 not significant df; † $p < 0.10$; *=$p < 0.05$, **=$p < 0.01$, ***=$p < 0.001$

Regarding grade level, 105 studies took place with students in grades 3–5, 50 studies included students in grades K-2, and 66 studies included students in grades 6–12. Holding all other covariates constant, grade level did not significantly describe the heterogeneity in our sample. Specifically, it appears that interventions for students in grades 3–5 are as efficacious as interventions for students in grades K-2 (full model $\beta = -1.67$, $p = 0.14$; reduced model $\beta = -0.05$, $p = 0.93$) or grades 6–12 (full model $\beta = 0.75$, $p = 0.40$; reduced model $\beta = -0.15$, $p = 0.13$).

Interventions were most often conducted in small groups ($k = 94$), closely followed by individually ($k = 88$). Groups of six or more occurred in 32 studies and 7 studies did not report or included mixed grouping. Intervention group size did not significantly describe any heterogeneity across either model. Specifically, interventions in small groups (2–5 students) were as efficacious as individual interventions (full model $\beta = -0.05$, $p = 0.93$; reduced model $\beta = 0.00$, $p = 0.93$), whole group interventions (6 or more students; full model $\beta = -2.27$, $p = 0.11$; reduced model $\beta = -0.10$, $p = 0.37$), or mixed/non-described studies (full model $\beta = 2.03$, $p = 0.41$; reduced model $\beta = 0.25$, $p = 0.42$) when holding all other covariates constant.

We examined heterogeneity in effects across content areas including algebra ($k = 14$), early numeracy ($k = 34$), fractions ($k = 33$), geometry ($k = 3$), measurement ($k = 3$), mixed problem types ($k = 13$), operations ($k = 60$), and word problems ($k = 61$). While again, the full model did not significantly influence heterogeneity in the full model, relative to fractions, the reduced model indicated significant heterogeneity in interventions focused on operations ($\beta = -0.34$, $p < 0.05$), problem-solving ($\beta = 0.33$, $p < 0.01$), and a near significant finding for early numeracy ($\beta = -0.43$, $p = 0.06$), suggesting that fraction interventions from our sample have resulted in significantly higher effects than interventions in these areas.

Dosage, categorized into four bands, included interventions less than 540-min ($k = 98$), between 540 and 1020 min ($k = 58$), between 1021 and 1350 min ($k = 19$), and interventions of 1350 min or more ($k = 46$). Compared to interventions of 1350 min or more, dosage did not significantly describe the heterogeneity in either model: less than 540 min (full model $\beta = 0.50$, $p = 0.58$; reduced model $\beta = -0.18$, $p = 0.09$); 540–1020 min (full model $\beta = 0.73$, $p = 0.48$; reduced model $\beta = -0.11$, $p = 0.19$); 1021–1350 min (full model $\beta = -2.03$, $p = 0.16$; reduced model $\beta = -0.16$, $p = 0.11$).

In the same model, we examined the impact of study characteristics. Study design was categorized as RCTs ($k = 119$), QED ($k = 16$), or SCRDs ($k = 86$). While study design comparing RCTs to QED ($\beta = 0.24$, $p = 0.85$) and SCRDs ($\beta = 0.51$, $p = 0.60$) was not significant in the full model, SCRD resulted in significantly higher effects than RCTs in the reduced model ($\beta = 0.45$, $p < 0.001$).

Most outcome measures were developed by the research team ($n = 529$), though 289 outcomes were reported from standardized mathematical measures. While the full model did not reveal a significant effect ($\beta = -0.16$, $p = 0.78$), the reduced model detected that standardized assessments resulted in significantly lower effects than researcher-developed measures ($\beta = -0.13$, $p < 0.05$).

For implementation fidelity, we examined studies that reported fidelity data related to dosage, adherence, and frequency ($k = 48$) compared to studies that

reported none ($k=47$), one ($k=43$), or two ($k=83$) of those categories, as outlined by the CEC Quality Indicators (Cook et al., 2014). Again, we found no significant variability related to whether studies reported implementation fidelity for none reported (full model $\beta=0.42$, $p=0.37$; reduced model $\beta=0.33$, $p=0.15$); one reported (full model $\beta=6.85$, $p=0.13$; reduced model $\beta=-0.04$, $p=0.67$); or two reported (full model $\beta=-1.63$, $p=0.25$; reduced model $\beta=-0.05$, $p=0.58$).

We also examined the impact of publication year, comparing studies published in 2005–2012 ($k=54$) to studies published in 2013 or later ($k=167$). We did not detect any significant moderation based on publication year (full model $\beta=0.57$, $p=0.40$; reduced model $\beta=0.10$, $p=0.31$).

Finally, for overall study quality, we calculated a quality percentage based on the CEC Quality Indicators (Cook et al., 2014). Based on their recommendations, we coded 22 to 24 indicators (depending on study design) to determine an overall quality score for each study. Then, we examined quality in three bands: studies scoring less than 75% ($k=58$), studies scoring between 75 and 90% ($k=98$), and studies that met greater than 90% ($k=65$) of the indicators. Findings did not demonstrate any significant heterogeneity based on overall study quality for studies meeting less than 75% of indicators (full model $\beta=-1.33$, $p=0.25$; reduced model $\beta=-0.09$, $p=0.46$) or studies meeting 75–90% of indicators (full model $\beta=0.39$, $p=0.55$; reduced model $\beta=-0.03$, $p=0.73$).

Discussion

This meta-analysis examined the effects of mathematics interventions for students with MD and explored potential moderators of these effects. The results underscore the significant benefits of implementing such interventions in schools. Specifically, when including outliers, a robust effect size (Hedges' g) of 1.29 was observed. This magnitude is notably larger than those identified in previous meta-analyses (e.g., Dennis et al., 2016, $g=0.52$; Myers et al., 2021, $g=0.52$; Stevens et al., 2018, $g=0.85$). Certain distinct features of our meta-analysis, such as the inclusion of SCRD and a broader range of studies, may have influenced this larger observed effect size. When excluding outliers, our findings ($g=0.56$) closely resembled those from earlier meta-analyses. As per Bloom et al. (2008), the average yearly effect size for student achievement in mathematics spans from 1.14 in kindergarten to 0.01 in grade 12. Thus, an effect size of 0.56 can be interpreted as providing roughly half a year of mathematical progress, despite the interventions in this study lasting an average of 17.4 h.

Intervention Characteristics

Delving deeper, our meta-regression evaluated the heterogeneity within effect sizes based on intervention characteristics, including grade level, group size, content area, and dosage. Contrary to Dennis et al. (2016), grade level did not emerge as a significant moderator in our findings, perhaps attributable to the inclusion of secondary

settings in our analysis. Previous research has indicated the importance of early mathematics intervention, particularly in the early elementary years due to the suggestion of larger effect sizes (Gersten et al., 2009a, 2009b). However, this meta-analysis' finding that there's not a significant difference in effect size between kindergarten through grade 2 and grade 3 through grade 5 or grade 6 through 12 supports the idea that interventions can be effective across grade levels.

In our study, group size was not a significant moderator influencing outcomes. Although interventions with whole groups showed a slightly negative effect compared to small groups in our full model ($\beta = -2.27$), this difference was not statistically significant. Furthermore, this disparity diminished ($\beta = -0.10$) when outliers were removed from the analysis. Our findings resonate with the meta-analyses of Lein et al. (2020) and Myers et al. (2023), who also reported no substantial impact of group size on results. This is in line with applied research by Clarke et al., (2017, 2021) and Doabler et al. (2019), where randomizing students with (MD) into different group sizes did not yield significant variations in math outcomes. Reflecting on Hattie's influential 2008 study, which found a modest effect of small group instruction on student achievement (effect size = 0.13), he recommended emphasizing factors like feedback, instructional quality, and teaching strategies over group size. The results of our meta-analysis reinforce this perspective, suggesting that schools can efficiently allocate resources for interventions for students with MD without focusing on group size.

For mathematics content area, our analysis revealed that students with MD made smaller gains in operations ($\beta = -0.34$) and problem-solving ($\beta = -0.33$) compared to fractions. Interventions focused on early numeracy were approaching significance ($\beta = -0.43$). These findings corroborate those from Stevens et al. (2018) that fraction interventions yielded larger outcomes that those focusing on problem-solving and operations. A plausible hypothesis for why interventions focused on fractions yield higher effect sizes may be rooted in the intricate nature of fractions and the expertise of research groups devoted to this domain. Fractions are inherently more complex than other foundational mathematics concepts, requiring students to shift from the discrete understanding of numbers to grasping their continuous and relational nature (Siegler et al., 2013). This added complexity might mean that effective intervention methods are more noticeable, and students make much greater gains than business as usual instruction. It is also worth noting that distinguished research groups, particularly those led by Lynn Fuchs, have invested significant efforts into understanding and teaching fractions. The concentrated efforts and expertise of those groups have resulted in well-structured, evidence-based interventions that cater specifically to students with MD (e.g., Fuchs et al., 2013, 2016). Thus, it may behoove schools and teachers to reference the instructional components of these evidence-based interventions when deciding what to include within their mathematics interventions.

Intervention duration in this meta-analysis showed a trend toward a negative effect size for shorter interventions (i.e., less than 540 min [9 h]; $\beta = -0.18$) over longer interventions (i.e., greater than 1350 min [22.5 h]) approaching significance ($p = 0.09$) in the model excluding outliers. There were no differences in interventions lasting between 541–1020 and 1020–1350 min compared to longer interventions

lasting more than 1350 min. In the special education literature, dosage is commonly recommended as a means of intensifying interventions for students who have not adequately responded to high-quality instruction (Powell et al., 2022). However, recommending duration as a means for intensification has its limitations. While the shortest interventions (i.e., those that lasted less than 9 h) may have resulted in smaller effect sizes, there were no discernible differences between interventions that lasted more than 9 h from those that lasted over 22.5 h.

Study Characteristics

Next, we explored the potential moderating effect of study characteristics, including research design, type of outcome measure, implementation fidelity, study year, and study quality. In the model without outliers, studies employing SCRD had significantly higher effect sizes ($\beta = 0.45$) than those using RCTs. It is possible that characteristics inherent to SCRD, such as flexibility to adapt to individual response (Kazdin, 2011), may contribute to stronger effect sizes. Another plausible explanation is the lack of psychometrically valid assessments used in SCRD, which could inflate effect sizes (Peltier et al., 2021).

Outcomes based on type of measure were also examined as potential moderators of the effect size. Standardized measures were statistically significantly lower ($\beta = -0.13$) than researcher-developed measures. Vaughn and Swanson (2015) noted that standardized measures are less likely to be aligned with intervention content, making it more challenging to observe changes at posttest. For instance, Myers et al. (2021) found that secondary students scored significantly lower on standardized measures after mathematics interventions ($\beta = -0.50$; p = 0.04) compared to researcher-developed measures, while Jitendra et al.'s (2021) reported that students receiving tier two interventions scored lower on standardized measures ($\beta = -0.27$; $p < 0.05$) than researcher developed measures.

Concerning implementation fidelity, we evaluated how fidelity was reported by researchers and whether there was a moderating effect based on that approach. We categorized each study based on the presence of one, two, or three of the CEC (Cook et al., 2015) criteria for assessing treatment fidelity. These criteria include whether the studies reported: (1) dosage, (2) adherence to the intervention, and (3) regularity of tracking either dosage or adherence throughout the study. Our meta-analysis revealed that effect sizes did not differ based on whether the researcher reported zero, one, two, or all three criteria. These results may suggest that the inherent effectiveness of the interventions are robust to variations in implementation fidelity, or that current fidelity measures may not capture the subtle nuances that significantly impact outcomes.

Finally, we considered the potential moderating effect of year and study quality on the effectiveness of mathematics interventions. Our findings suggest that intervention effects did not differ based on either of these two moderators. Studies conducted after 2013 (with outliers, $\beta = 0.57$; without outliers, $\beta = 0.10$) were as efficacious as those conducted prior to 2013, before common state standards had been established. Researchers, such as Williams et al. (2022), have posited that the counterfactual (i.e.,

typical classroom instruction or business-as-usual) has improved due to improved instructional practices or adoption of innovative programs in schools. If that is the case, our results suggest that intervention research has equally improved over time. In addition, there were no differences in effect sizes based on the quality of the study. That is, effect sizes did not vary from those that might be considered of lesser rigor to those of higher methodological quality.

Impact of Outliers

Given the large difference in effects between the models with and without outliers, we explored possible trends in outlier effects. To first understand this influence, consider the differences in heterogeneity across the samples. For the full model, the between study variance (τ^2) was 1.80. For the model excluding outliers, the τ^2 was reduced to 0.34. Such a change signals that the full model includes significant variation across studies, which is likely to influence the overall impact of some studies, an underlying reason for why we found it important to include both models. Furthermore, results from our meta-regressions revealed differences in the following moderators: (a) mathematical content area, (b) research design, and (c) measure standardization. In each of these cases, the meta-regression that excluded major outlier effects revealed significant covariates. These findings indicate that there may be study-level patterns in our outlier data.

To explore these patterns, we inspected our raw data for each moderator. First, regarding mathematical content, an examination of our data demonstrates that a disproportionate number of problem-solving interventions had inflated effects (35.9%, relative to the 27.4% of studies focused on problem-solving included in the full sample). Furthermore, all studies that focused on geometry had inflated effects that were excluded from the reduced model. Second, regarding research design, we found that a disproportionate number of effects from SCRDs were outliers. Specifically, of the 39.5% of effects from SCRDs in our full model, 45.7% of outliers implemented SCRDs. Furthermore, extremely large effects (i.e., studies with effects greater than $g = 4.0$) were more likely to be from SCRDs (70.4%). Also of interest, all inflated negative effects ($g < -1.2$) were from group designs (RCTs or QED). Lastly, we found no pattern in the outliers regarding whether the measure was researcher-developed.

Together, these data suggest that both research design and specific content area may be related to overall effect size. While content area data is difficult to interpret, save for few studies have focused on geometry skills specifically, representing an area of continued need for research, we believe that the findings related to research design may have specific implications for future studies. Specifically, this meta-analysis represents a novel, yet growing, approach to meta-analyses by including both group and SCRDs within the same meta-regression. While we argue this approach has many benefits to the field of mathematics intervention research, we hypothesize that this finding, both of inflated and larger effects, may replicate into other fields of education, such as reading or behavior. While as a construct, the BC-SMD is theoretically the same as the Hedges' g (Chen et al., 2023); BC-SMD data can be more

sensitive to the measures used and participant samples, given the small sample size. Thus, the field is in need of more interpretable benchmark effects to fully understand the impact of SCRD relative to group design research in mathematical interventions.

Limitations

Our results should be interpreted in light of several limitations. First, in the current meta-analysis, we used a broad term for defining MD, and the participants were grouped based on author-specified criteria for being MD or at-risk for MD. This approach limits the ability to determine the effectiveness of the intervention for specific groups of students. Nevertheless, employing a broad definition, especially considering the number of students who may be experiencing MD due to the impact of COVID-19, allows for a comprehensive analysis of the overall impact of interventions, providing valuable understanding of their efficacy across a larger range of students.

Second, future research should further explore the effectiveness of interventions wherein the content area of the interventions is aligned with the participants' grade level curricular standards. Although the current study analyzed grade-level as a moderator for both elementary and secondary, the study did not evaluate if interventions were more or less effective when aligned with the curriculum and sequence of the state mathematics standards for the grade level. Understanding the impact of curricular alignment may provide important insight for the successful development and implementation of mathematics intervention studies, as has been recommended and identified as a need in intervention research (Jitendra et al., 2013).

Third, while our analysis includes various potential moderators, it is important to clarify that the statistical insignificance of certain factors like dosage level or group size does not denote an absence of effect. Rather, such findings often indicate inconclusive results, attributable to limitations such as a smaller number of studies or methodological differences (Gurevitch et al., 2018). This distinction is vital for accurately interpreting the impact of these moderators on educational outcomes.

Implications for Future Research and Practice

The results of this meta-analysis draw attention to several implications for future research. First, to develop a comprehensive picture of the impact of mathematics interventions for students with MD, the evidence-base must reflect a robust understanding of all mathematics content areas. Unfortunately, due to the limited number of geometry studies ($k=4$), along with the large effect sizes, geometry studies were excluded from the analysis without outliers. Thus, the results of the model without outliers reflect a portion of extant research on mathematics interventions for students with MD and cannot be viewed as all encompassing. There were also too few measurement studies ($k=3$) to have reliable outcomes in either model. Additional empirical research that evaluates geometry and measurement outcomes for students with MD is warranted to ensure comprehensive evaluations of the evidence base are possible.

Second, it is imperative for researchers to focus on developing effective multi-component interventions for other critical mathematics domains such as problem solving and early numeracy. This expansion is as crucial as the progress made in interventions for fractions. By doing so, we ensure that we are crafting robust interventions that can significantly influence the trajectory of mathematical learning and outcomes for students with MD, setting a strong foundation for their future academic achievements in mathematics.

Third, the development of valid and reliable measures for SCRD studies is necessary to improve the trustworthiness of results of studies that utilize this type of research design. Although our findings suggest that SCRD studies led to higher mathematics outcomes, an evaluation of results with and without outliers suggests possible effect size inflation. One possible reason for inflation may be the limited psychometric evidence available for selected measures. Thus, it is necessary for researchers to continue to develop reliable and valid assessment measures that can be used in SCRD.

In addition to implications for future research, it is necessary to address implications for practice related to implementation of mathematics interventions for students with MD. The findings of this study underscore the necessity of incorporating research-based interventions, such as "Fraction Face-Off" (e.g., Fuchs et al., 2013), in the mathematics curriculum for students with MD. These interventions have shown to significantly improve student performance focused on critical foundational fraction concepts essential for their success in more advanced stages of education like middle and high school (Siegler et al., 2013). Moreover, implementing the evidence-based practices found within these effective interventions, which include systematic instruction and the use of multiple, visual representations (Fuchs et al., 2021), can be beneficial not only for teaching fractions but also when applied to other mathematical content areas, such as early numeracy.

Lastly, factors such as grade level and group size did not have discernable effects on mathematics outcomes. Consequently, it is advisable for teachers to prioritize employing effective instructional strategies, such as providing systematic instruction, using clear and concise mathematics language, and incorporating timed activities (See Fuchs et al., 2021) over decisions about group size for mastery of foundational skills associated with mathematics proficiency at all grade levels. The concept of diminishing returns was evident in relation to instructional minutes, highlighting the importance of balancing sufficient time for content mastery with the efficiency of instruction. It appears that allocating over 9 h for instruction or until mastery is achieved might be optimal. Beyond this point, the benefits of additional instruction on the same content start to wane. Instead, it may be more advantageous to progress to new content, revisiting previous concepts through instructional review (Gersten et al., 2009a, 2009b).

Conclusion

In conclusion, this meta-analysis of mathematics interventions for students with MD provides valuable insights into the literature prior to the COVID-19 pandemic, offering a solid foundation for implementing effective interventions in schools to address the learning disruptions experienced during the most recent school years. The results

demonstrate that interventions targeting mathematics can yield significant positive effects on students' mathematics outcomes, offering a glimmer of hope. As schools navigate the challenges brought on by the COVID-19 pandemic, it is essential to adapt and refine interventions to meet the needs of the students and to target content that builds pre-algebraic readiness. Continuous collaborative efforts by researchers and teachers to strengthen students' mathematics proficiency, aiming not only to meet but surpass pre-COVID-19 pandemic levels, will help build a brighter future for all.

Supplementary Information   The online version contains supplementary material available at https://doi.org/10.1007/s10648-023-09843-0.

Data Availability   The dataset and analysis are available in the supplementary materials.

Declarations

Conflict of Interest   The authors declare no competing interests.

References

Andersson, U. (2008). Mathematical competencies in children with different types of learning difficulties. *Journal of Educational Psychology, 100*(1), 48–66. https://doi-org.ezproxy.lib.utexas.edu/10.1037/0022-0663.100.1.48.

Aunio, P., Heiskari, P., Van Luit, J. E., & Vuorio, J. M. (2015). The development of early numeracy skills in Kindergarten in low-, average-and high-performance groups. *Journal of Early Childhood Research, 13*(1), 3–16. https://doi.org/10.1177/1476718X14538722

Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014). State and trait effects on individual differences in children's mathematical development. *Psychological Science, 25*(11), 2017–2026.

Baroody, A. J., & Lai, M. (2022). The development and assessment of counting-based cardinal number concepts. *Educational Studies in Mathematics, 111*(2), 185–205.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness, 1*(4), 289–328. https://doi.org/10.1080/19345740802400072

Booth, J. L., & Newton, K. J. (2012). Fractions: Could they really be the gatekeeper's doorman? *Contemporary Educational Psychology, 37*(4), 247–253. https://doi.org/10.1016/j.cedpsych.2012.07.001

Borenstein, M., Cooper, H., Hedges, L., & Valentine, J. (2009). Effect sizes for continuous data. *The Handbook of Research Synthesis and Meta-Analysis, 2*, 221–235.

Bouck, E. C., Satsangi, R., & Park, J. (2018). The concrete–representational–abstract approach for students with learning disabilities: An evidence-based practice synthesis. *Remedial and Special Education, 39*(4), 211–228. https://doi-org.ezproxy.lib.utexas.edu/10.1177/0741932517721712.

Bryant, D. P., Pfannenstiel, K. H., Bryant, B. R., Roberts, G., Fall, A. M., Nozari, M., & Lee, J. (2021). Improving the mathematics performance of second-grade students with mathematics difficulties through an early numeracy intervention. *Behavior Modification, 45*(1), 99–121. https://doi-org.ezproxy.lib.utexas.edu/10.1177/0145445519873651

Chen, J., Li, L., & Zhang, D. (2021). Students with specific difficulties in geometry: Exploring the TIMSS 2011 data with plausible values and latent profile analysis. *Learning Disability Quarterly, 44*(1), 11–22. https://doi.org/10.1177/0731948720905099

Chen, M., Pustejovsky, J. E., Klingbeil, D. A., & Van Norman, E. R. (2023). Standardized mean differences: Flexible methods for single-case designs. *Journal of School Psychology, 98*, 16–38. https://doi.org/10.1016/j.jsp.2023.02.002

Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283–292. https://doi-org.ezproxy.lib.utexas.edu/10.3102/0013189X16656615

Chodura, S., Kuhn, J. T., & Holling, H. (2015). Interventions for children with mathematical difficulties. *Zeitschrift für Psychologie, 223*, 129–144. https://doi-org.ezproxy.lib.utexas.edu/10.1027/2151-2604/a000211

Chong, S. L., & Siegel, L. S. (2008). Stability of computational deficits in math learning disability from second through fifth grades. *Developmental Neuropsychology, 33*(3), 300–317.

Chu, F. W., VanMarle, K., & Geary, D. C. (2013). Quantitative deficits of preschool children at risk for mathematical learning disability. *Frontiers in Psychology, 4*, 195.

Cirino, P. T., Fuchs, L. S., Elias, J. T., Powell, S. R., & Schumacher, R. F. (2015). Cognitive and mathematical profiles for different forms of learning difficulties. *Journal of Learning Disabilities, 48*(2), 156–175.

Cirino, P. T., Tolar, T. D., & Fuchs, L. S. (2019). Longitudinal algebra prediction for early versus later takers. *The Journal of Educational Research, 112*(2), 179–191. https://doi-org.ezproxy.lib.utexas.edu/10.1080/00220671.2018.1486279

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

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. Crossref.

Clarke, D. M., Stephens, M., & Sullivan, P. (2021). Refections on themes in professional contributions by David Clarke. *ZDM*. https://doi.org/10.1007/s11858-021-01288-5

Clements, D., & Sarama, J. (2013). Math in the early years: A strong predictor for later school success. *ECS Research Brief, the Progress of Educational Reform, 4*(5), 1–7.

Clements, D. H., Sarama, J., Baroody, A. J., & Joswick, C. (2020). Efficacy of a learning trajectory approach compared to a teach-to-target approach for addition and subtraction. *ZDM*. https://doi.org/10.1007/s11858-019-01122-z

Codding, R. S., Burns, M. K., & Lukito, G. (2011). Meta-analysis of mathematic basic-fact fluency interventions: A component analysis. *Learning Disabilities Research & Practice, 26*(1), 36–47. https://doi-org.ezproxy.lib.utexas.edu/10.1111/j.1540-5826.2010.00323.x

Cook, B., Buysse, V., Klingner, J., Landrum, T., McWilliam, R., Tankersley, M., & Test, D. (2014). Council for Exceptional Children: Standards for evidence-based practices in special education. *Teaching Exceptional Children, 46*(6), 206.

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education, 36*(4), 220–234. https://doi-org.ezproxy.lib.utexas.edu/10.1177/0741932514557271

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

Cowan, R., & Powell, D. (2014). The contributions of domain-general and numerical factors to third-grade arithmetic skills and mathematical learning disability. *Journal of Educational Psychology, 106*(1), 214.

Dennis, M. S., Sharp, E., Chovanes, J., Thomas, A., Burns, R. M., Custer, B., & Park, J. (2016). A meta-analysis of empirical research on teaching students with mathematics learning difficulties. *Learning Disabilities Research & Practice, 31*(3), 156–168. https://doi-org.ezproxy.lib.utexas.edu/10.1111/ldrp.12107

Dessemontet, R. S., Martinet, C., de Chambrier, A. F., Martini-Willemin, B. M., & Audrin, C. (2019). A meta-analysis on the effectiveness of phonics instruction for teaching decoding skills to students with intellectual disability. *Educational Research Review, 26*, 52–70.

Doabler, C. T., Clarke, B., Kosty, D., Turtura, J. E., Firestone, A. R., Smolkowski, K., Jungjohann, K., Brafford, T. L., Nelson, N. J., Sutherland, M., Fien, H., & Maddox, S. A. (2019). Efficacy of a first-grade mathematics intervention on measurement and data analysis. *Exceptional Children, 86*(1), 77–94.

Doabler, C. T., Clarke, B., Kosty, D., Sutherland, M., Turtura, J. E., Firestone, A. R., Kimmel, G. L., Brott, P., Brafford, T. L., Nelson Fien, N. J., Smolkowski, K., & Jungjohann, K. (2022). Promoting understanding of measurement and statistical investigation among second-grade students with mathematics difficulties. *Journal of Educational Psychology, 114*(3), 560.

Doane, D. P., & Seward, L. E. (2011). Measuring skewness: A forgotten statistic? *Journal of Statistics Education*, *19*(2). https://doi-org.ezproxy.lib.utexas.edu/10.1080/10691898.2011.11889611

Dyson, N. I., Jordan, N. C., & Glutting, J. (2011). Number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities, 46*(2), 166–181. https://doi.org/10.1177/0022219411410233

Ennis, R. P., & Losinski, M. (2019). Interventions to improve fraction skills for students with disabilities: A meta-analysis. *Exceptional Children, 85*(3), 367–386.

Fisher, Z., & Tipton, E. (2015). robumeta: An R-package for robust variance estimation in meta-analysis. *arXiv preprint arXiv:1503.02220*.

Flores, M. M., Hinton, V., & Strozier, S. D. (2014). Teaching subtraction and multiplication with regrouping using the concrete-representational-abstract sequence and strategic instruction model. *Learning Disabilities Research & Practice, 29*(2), 75–88. https://doi-org.ezproxy.lib.utexas.edu/10.1111/ldrp.12032

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide (NCEE 2014–4005)*. Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved March 1, 2023, from the NCEE website: http://whatworks.ed.gov

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., Hamlett, C. L., & Zumeta, R. O. (2009). Remediating number combination and word problem deficits among students with mathematics difficulties: A randomized control trial. *Journal of Educational Psychology, 101*(3), 561.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2010). The effects of strategic counting instruction with and without deliberate practice on number combination skill among students with mathematics difficulties. *Learning and Individual Differences, 20*(2), 89–100.

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., Jordan, N. C., Siegler, R., Gersten, R., & Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology, 105*(3), 683.

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Malone, A. S., Wang, A., Hamlett, C. L., Jordan, N. C., Siegler, R. S., & Changas, P. (2016). Effects of intervention to improve at-risk fourth graders' understanding, calculations, and word problems with fractions. *The Elementary School Journal, 116*(4), 625–651.

Fuchs, L. S., Fuchs, D., & Malone, A. S. (2017). The taxonomy of intervention intensity. *Teaching Exceptional Children, 50*(1), 35–43.

Fuchs, L. S., Newman-Gonchar, R., Schumacher, R., Dougherty, B., Bucka, N., Karp, K. S., Woodward, J., Clarke, B., Jordan, N. C., Gersten, R., Jayanthi, M., Keating, B., & Morgan, S. (2021). *Assisting students struggling with mathematics: Intervention in the elementary grades (WWC 2021006)*. Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved February 25, 2023 from http://whatworks.ed.gov/

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PloS one, 8*(1), e54651. https://doi.org/10.1371/journal.pone.0054651

Gersten, R., Chard, D., & Baker, S. (2000). Factors enhancing sustained use of research-based instructional practices. *Journal of Learning Disabilities, 33*(5), 445–456.

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

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005b). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 293–304. https://doi.org/10.1177/00222194050380040301

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009a). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools (NCEE 2009–4060)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved February 2, 2023 from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009b). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components.

Review of Educational Research, 79(3), 1202–1242. https://doi-org.ezproxy.lib.utexas.edu/10.3102/0034654309334431

Gonzales, P., Guzmán, J. C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., & Williams, T. (2004). *Highlights From the Trends in International Mathematics and Science Study (TIMSS) 2003* (NCES 2005–005). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office. Retrieved on February 5, 2023, from https://nces.ed.gov/pubs2005/2005005.pdf

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature, 555*, 175–182. https://doi.org/10.1038/nature25753

Hannula-Sormunen, M. M., Lehtinen, E., & Räsänen, P. (2015). Preschool children's spontaneous focusing on numerosity, subitizing, and counting skills as predictors of their mathematical performance seven years later at school. *Mathematical Thinking and Learning, 17*(2–3), 155–177. https://doi.org/10.1080/10986065.2015.1016814

Hansen, N., Jordan, N. C., Fernandez, E., Siegler, R. S., Fuchs, L., Gersten, R., & Micklos, D. (2015). General and math-specific predictors of sixth-graders' knowledge of fractions. *Cognitive Development, 35*, 34–49. https://doi.org/10.1016/j.cogdev.2015.02.001

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Routledge.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Academic Press.

Higgins, J. P., & Green, S. (Eds.). (2008). *Cochrane handbook for systematic reviews of interventions.* Wiley Online Library. https://doi.org/10.1002/9780470712184

Hill, H. C., & Erickson, A. (2019). Using implementation fidelity to aid in interpreting program impacts: A brief review. *Educational Researcher, 48*(9), 590–598. https://doi.org/10.3102/0013189X19891436

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*(2), 165–179. https://doi-org.ezproxy.lib.utexas.edu/10.1177/001440290507100203

Jitendra, A. K., Rodriguez, M., Kanive, R., Huang, J. P., Church, C., Corroy, K. A., & Zaslofsky, A. (2013). Impact of small-group tutoring interventions on the mathematical problem solving and achievement of third-grade students with mathematics difficulties. *Learning Disability Quarterly, 36*(1), 21–35. https://doi.org/10.1177/073194871245

Jitendra, A. K., Harwell, M. R., Dupuis, D. N., & Karl, S. R. (2017). A randomized trial of the effects of schema-based instruction on proportional problem-solving for students with mathematics problem-solving difficulties. *Journal of Learning Disabilities, 50*(3), 322–336.

Jitendra, A. K., Lein, A. E., Im, S. H., Alghamdi, A. A., Hefte, S. B., & Mouanoutoua, J. (2018). Mathematical interventions for secondary students with learning disabilities and mathematics difficulties: A meta-analysis. *Exceptional Children, 84*(2), 177–196. https://doi-org.ezproxy.lib.utexas.edu/10.1177/0014402917737467

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

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology, 45*(3), 850–867. https://doi.org/10.1037/a0014939

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B., & Irwin, C. (2012). Building kindergartners' number sense: A randomized controlled study. *Journal of Educational Psychology, 104*(3), 647–660. https://doi.org/10.1037/a0029018

Jordan, N. C., Rodrigues, J., Hansen, N., & Resnick, I. (2017). Fraction development in children: Importance of building numerical magnitude understanding. In *Acquisition of complex arithmetic skills and higher-order mathematics concepts* (pp. 125–140). Academic Press. https://doi.org/10.1016/B978-0-12-805086-6.00006-0

Kazdin, A. E. (2011). Evidence-based treatment research: Advances, limitations, and next steps. *American Psychologist, 66*(8), 685.

Kim, J., Kim, J., DesJardins, S. L., & McCall, B. P. (2015). Completing Algebra II in high school: Does it increase college access and success? *The Journal of Higher Education, 86*(4), 628–662. https://doi.org/10.1353/khe.2015.0018

Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Educational Psychology, 105*(1), 162.

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher, 49*(4), 241–253.

Kroesbergen, E. H., & Van Luit, J. E. (2003). Mathematics interventions for children with special educational needs: A meta-analysis. *Remedial and Special Education, 24*(2), 97–114.

Kroesbergen, E. H., Van Luit, J. E., & Naglieri, J. A. (2003). Mathematical learning difficulties and PASS cognitive processes. *Journal of Learning Disabilities, 36*(6), 574–582.

Lambert, K., & Moeller, K. (2019). Place-value computation in children with mathematics difficulties. *Journal of Experimental Child Psychology, 178*, 214–225.

Lee, I., Grover, S., Martin, F., Pillai, S., & Malyn-Smith, J. (2020). Computational thinking from a disciplinary perspective: Integrating computational thinking in K-12 science, technology, engineering, and mathematics education. *Journal of Science Education and Technology, 29*, 1–8.

Lei, Q., Mason, R. A., Xin, Y. P., Davis, J. L., David, M., & Lory, C. (2020). A meta-analysis of single-case research on mathematics word problem-solving interventions for English learners with learning disabilities and mathematics difficulties. *Learning Disabilities Research & Practice, 35*(4), 201–217. https://doi-org.ezproxy.lib.utexas.edu/10.1111/ldrp.12233

Lein, A. E., Jitendra, A. K., & Harwell, M. R. (2020). Effectiveness of mathematical word problem solving interventions for students with learning disabilities and/or mathematics difficulties: A meta-analysis. *Journal of Educational Psychology, 112*(7), 1388–1408. https://doi-org.ezproxy.lib.utexas.edu/10.1037/edu0000453

Li, Y., Wang, K., Xiao, Y., & Froyd, J. E. (2020). Research and trends in STEM education: A systematic review of journal publications. *International Journal of STEM Education, 7*(1), 1–16.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms (NCSER 2013–3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, US Department of Education.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*(12), 1181–1209.

Liu, D., Mao, Y., Cai, W., Lei, Q., Kang, R., & Zeng, Y. (2023). Meta-analysis of tablet-mediated interventions to teach mathematics for individuals with autism spectrum disorder and/or intellectual disability. *Journal of Special Education Technology, 0*(0). https://doi-org.ezproxy.lib.utexas.edu/10.1177/01626434231180579

Losinski, M., Cuenca-Carlino, Y., Zablocki, M., & Teagarden, J. (2014). Examining the efficacy of self-regulated strategy development for students with emotional or behavioral disorders: A meta-analysis. *Behavioral Disorders, 40*(1), 52–67.

Maggin, D. M., Cook, B. G., & Cook, L. (2018). Using single-case research designs to examine the effects of interventions in special education. *Learning Disabilities Research & Practice, 33*(4), 182–191.

Mancl, D. B., Miller, S. P., & Kennedy, M. (2012). Using the concrete-representational-abstract sequence with integrated strategy instruction to teach subtraction with regrouping to students with learning disabilities. *Learning Disabilities Research & Practice, 27*(4), 152–166.

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE, 6*(9), e23749.

Missall, K. N., Mercer, S. H., Martínez, R. S., & Casebeer, D. (2012). Concurrent and longitudinal patterns and trends in performance on early numeracy curriculum-based measures in kindergarten through third grade. *Assessment for Effective Intervention, 37*(2), 95–106. https://doi.org/10.1177/1534508411430322

Mistretta, R. M. (2000). Enhancing geometric reasoning. *Adolescence, 35*(138), 365.

Morgan, P. L., & Sideridis, G. D. (2006). Contrasting the effectiveness of fluency interventions for students with or at risk for learning disabilities: A multilevel random coefficient modeling meta-analysis. *Learning Disabilities Research & Practice, 21*(4), 191–210.

Morin, L. L., Watson, S. M., Hester, P., & Raver, S. (2017). The use of a bar model drawing to teach word problem solving to students with mathematics difficulties. *Learning Disability Quarterly, 40*(2), 91–104.

Myers, J. A., Brownell, M. T., Griffin, C. C., Hughes, E. M., Witzel, B. S., Gage, N. A., Peyton, D., Acosta, K., & Wang, J. (2021). Mathematics interventions for adolescents with mathematics difficulties: a meta-analysis. *Learning Disabilities Research & Practice, 36*(2), 145–166. https://doi-org.ezproxy.lib.utexas.edu/10.1111/ldrp.12244

Myers, J. A., Hughes, E. M., Witzel, B. S., Anderson, R. D., & Owens, J. (2023). A meta-analysis of mathematical interventions for increasing the word problem solving performance of upper elementary and secondary students with mathematics difficulties. *Journal of Research on Educational Effectiveness, 16*(1), 1–35. https://doi-org.ezproxy.lib.utexas.edu/10.1080/19345747.2022.2080131

Namkung, J. M., & Fuchs, L. S. (2016). Cognitive predictors of calculations and number line estimation with whole numbers and fractions among at-risk students. *Journal of Educational Psychology, 108*(2), 214.

National Center for Education Statistics. (2022). *National Assessment of educational Progress*. U.S. Department of Education. Retrieved March 2, 2023, from http://nces.ed.gov/nationsreportcard/

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. National Council of Teachers of Mathematics.

National Governors Association Center for Best Practices & Council of Chief State School Officers (2010). *Common Core state standards mathematics*. Washington, DC: National Governors Association Center for Best Practices & Council of Chief State School Officers.

Nelson, G., & Powell, S. R. (2018). A systematic review of longitudinal studies of mathematics difficulty. *Journal of Learning Disabilities, 51*(6), 523–539. https://doi-org.ezproxy.lib.utexas.edu/10.1177/0022219417714773

Nozari, M., Bryant, D. P., Doabler, C. T., Falcomata, T. S., Liu, M., & Grimes, R. (2021). Intensifying instruction to address treatment resistance to early mathematics interventions. *Current Psychology, 42*, 10355–10374. https://doi.org/10.1007/s12144-021-02320-3

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

Odden, T. O. B., Marin, A., & Rudolph, J. L. (2021). How has Science Education changed over the last 100 years? An analysis using natural language processing. *Science Education, 105*(4), 653–680.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T., Mulrow, C. D., & Moher, D. (2020). Mapping of reporting guidance for systematic reviews and meta-analyses generated a comprehensive item bank for future reporting guidelines. *Journal of Clinical Epidemiology, 118*, 60–68.

Peltier, C., Garwood, J. D., McKenna, J., Peltier, T., & Sendra, J. (2021). Using the SRSD instructional approach for argumentative writing: A look across the content areas. *Learning Disabilities Research & Practice, 36*(3), 224–234. https://doi-org.ezproxy.lib.utexas.edu/10.1111/ldrp.12255

Polanin, J. R., Pigott, T. D., Espelage, D. L., & Grotpeter, J. K. (2019). Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. *Research Synthesis Methods, 10*(3), 330–342.

Powell, S. R., & Fuchs, L. S. (2015). Intensive intervention in mathematics. *Learning Disabilities Research and Practice, 30*(4), 182–192. https://doi.org/10.1111/ldrp.12087

Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P. T., & Fletcher, J. M. (2009). Effects of fact retrieval tutoring on third-grade students with math difficulties with and without reading difficulties. *Learning Disabilities Research and Practice, 24*(1), 1–11. https://doi.org/10.1111/j.1540-5826.2008.01272.x

Powell, S. R., Doabler, C. T., Akinola, O. A., Therrien, W. J., Maddox, S. A., & Hess, K. E. (2020). A synthesis of elementary mathematics interventions: Comparisons of students with mathematics difficulty with and without comorbid reading difficulty. *Journal of Learning Disabilities, 53*(4), 244–276. https://doi.org/10.1177/0022219419881646

Powell, S. R., Benz, S. A., Mason, E. N., & Lembke, E. S. (2022). How to structure and intensify mathematics intervention. *Beyond Behavior, 31*(1), 5–15. https://doi.org/10.1177/10742956211072267

Protogerou, C., & Hagger, M. S. (2020). A checklist to assess the quality of survey studies in psychology. *Methods in Psychology, 3*(2020), 1–14. https://doi.org/10.1016/j.metip.2020.100031

Provasnik, S., Kastberg, D., Ferraro, D., Lemanski, N., Roey, S., & Jenkins, F. (2012). *Highlights From TIMSS 2011: Mathematics and Science Achievement of U.S. Fourth- and Eighth-Grade Students in an International Context* (NCES 2013-009). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved on February 8, 2023, from https://nces.ed.gov/pubs2013/2013009_1.pdf

Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology, 110*(4), 647–658.

Pustejovsky, J. E. (2018). Using response ratios for meta-analyzing single-case designs with behavioral outcomes. *Journal of School Psychology, 68*, 99–112.

Pustejovsky, J. E., & Rodgers, M. A. (2019). Testing for funnel plot asymmetry of standardized mean differences. *Research Synthesis Methods, 10*(1), 57–71.

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science, 23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

Pustejovsky, J. E., Chen, M., & Hamilton, B. (2021). Scdhlm: A web-based calculator for between-case standardized mean differences (version 0.5.2) Web application. Retrieved from https://jepusto.shinyapps.io/scdhlm

Raddatz, J., Kuhn, J. T., Holling, H., Moll, K., & Dobel, C. (2017). Comorbidity of arithmetic and reading disorder: Basic number processing and calculation in children with learning impairments. *Journal of Learning Disabilities, 50*(3), 298–308.

Raudenbush, S. W. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal, 45*(1), 206–230.

Reikerås, E. K. (2009). A comparison of performance in solving arithmetical word problems by children with different levels of achievement in mathematics and reading. *Investigations in Mathematics Learning, 1*(3), 49–72.

Rojas-Andrade, R., & Bahamondes, L. L. (2019). Is implementation fidelity important? A systematic review on school-based mental health programs. *Contemporary School Psychology, 23*(4), 339–350.

Rojo, M., King, S., Gersib, J., & Bryant, D. P. (2023). Rational number interventions for students with mathematics difficulties: A meta-analysis. *Remedial and Special Education, 44*(3), 225–238. https://doi.org/10.1177/07419325221105520

Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition, 102*(3), 361–395.

Rumrill P. D., Cook, B. G., & Stevenson, N. A. (Eds.). (2020). *Research in special education: Designs, methods, and applications* (3rd ed.). Charles C Thomas.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research learning trajectories for young children* (1st ed., Vol. 10). Routledge. https://doi.org/10.4324/9780203883785

Schulte, A. C., & Stevens, J. J. (2015). Once, sometimes, or always in special education: Mathematics growth and achievement gaps. *Exceptional Children, 81*(3), 370–387. https://doi.org/10.1177/0014402914563695

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology, 52*(2), 123–147. https://doi.org/10.1016/j.jsp.2013.11.005

Shanley, L., Clarke, B., Doabler, C. T., Kurtz-Nelson, E., & Fien, H. (2017). Early number skills gains and mathematics achievement: Intervening to establish successful early mathematics trajectories. *The Journal of Special Education, 51*(3), 177–188.

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., Susperreguy, M. I., & Chen, M. (2012). Early predictors of high school mathematics achievement. *Psychological Science, 23*(7), 691–697. https://doi-org.ezproxy.lib.utexas.edu/10.1177/0956797612440101

Siegler, R. S., Fazio, L. K., Bailey, D. H., & Zhou, X. (2013). Fractions: The new frontier for theories of numerical development. *Trends in Cognitive Sciences, 17*(1), 13–19.

Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis, 31*(4), 500–506.

Soland, J. (2013). Predicting high school graduation and college enrollment: Comparing early warning indicator data and teacher intuition. *Journal of Education for Students Placed at Risk, 18*(3–4), 233–262. https://doi.org/10.1080/10824669.2013.833047

Stevens, E. A., Rodgers, M. A., & Powell, S. R. (2018). Mathematics interventions for upper elementary and secondary students: A meta-analysis of research. *Remedial and Special Education, 39*(6), 327–340. https://doi-org.ezproxy.lib.utexas.edu/10.1177/0741932517731887

Stock, P., Desoete, A., & Roeyers, H. (2010). Detecting children with arithmetic disabilities from kindergarten: Evidence from a 3-year longitudinal study on the role of preparatory arithmetic abilities. *Journal of learning Disabilities, 43*(3), 250–268. https://doi.org/10.1177/0022219409345011

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2013). Intervention fidelity in special and general education research journals. *The Journal of Special Education, 47*(1), 3–13.

Talbott, E., Maggin, D. M., Van Acker, E. Y., & Kumm, S. (2018). Quality indicators for reviews of research in special education. *Exceptionality, 26*(4), 245–265.

Tian, J., & Siegler, R. S. (2017). Fractions learning in children with mathematics difficulties. *Journal of Learning Disabilities, 50*(6), 614–620.

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics, 40*(6), 604–634. https://doi-org.ezproxy.lib.utexas.edu/10.3102/1076998615606099

Toll, S. W., & Van Luit, J. E. (2012). Early numeracy intervention for low-performing kindergartners. *Journal of Early Intervention, 34*(4), 243–264. https://doi.org/10.1177/105381511347

Toste, J. R., Logan, J. A., Shogren, K. A., & Boyd, B. A. (2023). The next generation of quality indicators for group design research in special education. *Exceptional Children, 89*(4), 359–378. https://doi.org/10.1177/001440292211508

Tournaki, N., Bae, Y. S., & Kerekes, J. (2008). Rekenrek: A manipulative used to teach addition and subtraction to students with learning disabilities. *Learning Disabilities: A Contemporary Journal, 6*(2), 41–59.

Vaughn, S., & Swanson, E. A. (2015). Special education research advances knowledge in education. *Exceptional Children, 82*(1), 11–24. https://doi.org/10.1177/0014402915598781

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Vukovic, R. K., & Siegel, L. S. (2010). Academic and cognitive characteristics of persistent mathematics difficulty from first through fourth grade. *Learning Disabilities Research & Practice, 25*(1), 25–38.

Vukovic, R. K., Fuchs, L. S., Geary, D. C., Jordan, N. C., Gersten, R., & Siegler, R. S. (2014). Sources of individual differences in children's understanding of fractions. *Child Development, 85*(4), 1461–1476. https://doi.org/10.1111/cdev.12218

Wang, M. T., Ye, F., & Degol, J. L. (2017). Who chooses STEM careers? Using a relative cognitive strength and interest model to predict careers in science, technology, engineering, and mathematics. *Journal of Youth and Adolescence, 46*, 1805–1820.

Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational Psychology Review, 28*, 551–576.

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.

What Works Clearinghouse. (2020). *What Works Clearinghouse standards handbook, version 4.1.* National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved May 5, 2022, from https://ies.ed.gov/ncee/wwc/handbooks

Williams, R., Citkowicz, M., Miller, D. I., Lindsay, J., & Walters, K. (2022). Heterogeneity in mathematics intervention effects: Evidence from a meta-analysis of 191 randomized experiments. *Journal of Research on Educational Effectiveness, 15*(3), 584–634. https://doi-org.ezproxy.lib.utexas.edu/10.1080/19345747.2021.2009072

Wolf, B., & Harbatkin, E. (2022). Making sense of effect sizes: Systematic differences in intervention effect sizes by outcome measure type. *Journal of Research on Educational Effectiveness,* 1–28. https://doi.org/10.1080/19345747.2022.2071364

Zeneli, M., Thurston, A., & Roseth, C. (2016). The influence of experimental design on the magnitude of the effect size-peer tutoring for elementary, middle and high school settings: A meta-analysis. *International Journal of Educational Research, 76*, 211–223.

Zhang, D. (2021). Teaching geometry to students with learning disabilities: Introduction to the special series. *Learning Disability Quarterly, 44*(1), 4–10.

Zhang, D., & Xin, Y. P. (2012). A follow-up meta-analysis for word-problem-solving interventions for students with mathematics difficulties. *The Journal of Educational Research, 105*(5), 303–318. https://doi-org.ezproxy.lib.utexas.edu/10.1080/00220671.2011.627397

Authors and Affiliations

Megan Rojo[1] · **Jenna Gersib[2]** · **Sarah R. Powell[3]** · **Zhina Shen[4]** ·
Sarah G. King[2] · **Syeda Sharjina Akther[2]** · **Tessa L. Arsenault[2]** ·
Samantha E. Bos[5] · **Danielle O. Lariviere[2]** · **Xin Lin[6]**

✉   Megan Rojo
     megan.rojo@unt.edu

     Jenna Gersib
     jenna.gersib@utexas.edu

     Sarah R. Powell
     srpowell@utexas.edu

     Zhina Shen
     zns@utexas.edu

     Sarah G. King
     Sarah.gorsky@utexas.edu

     Syeda Sharjina Akther
     sharj07@utexas.edu

     Tessa L. Arsenault
     tarsenault@utexas.edu

     Samantha E. Bos
     the.samantha.bos@gmail.com

     Danielle O. Lariviere
     danielle.lariviere@utexas.edu

     Xin Lin
     xin.lin@utexas.edu

[1]   Educational Psychology, University of North Texas, 1300 W. Highland St., Denton, TX 76203,
      USA

[2]   Meadows Center for Preventing Educational Risk, The University of Texas at Austin, 1912
      Speedway, D4900, Austin, TX 78712, USA

[3]   The University of Texas at Austin, 1912 Speedway, D4900, Austin, TX 78712, USA

[4]   Innovative Learning Center, LLC., 1100 Peachtree Street NE, Atlanta 30309, Georgia

[5]   Gibson Consulting Group, 322 Blakeley Dr., San Antonio, TX 78792, USA

[6]   Faculty of Education, University of Macau, Avenida da Universidade, Taipa, Macau