Title
A Meta-Analysis of Single-Case Research on Mathematics Interventions for Students with Mathematics Learning Disabilities and Mathematics Difficulties

Author
Suchilt, Luisana

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

A Meta-Analysis of Single-Case Research on Mathematics Interventions for Students with Mathematics Learning Disabilities and Mathematics Difficulties

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Luisana Suchilt

September 2022

Oral Qualifying Exam Committee:
Dr. Asha K. Jitendra, Chairperson
Dr. Austin Johnson
Dr. Stephanie Moore

The Dissertation of Luisana Suchilt is approved:

Committee Chairperson

University of California, Riverside

Acknowledgements

I would like to thank my advisor, Dr. Asha Jitendra for her guidance, support, and wisdom throughout the dissertation process and my doctoral program. I am especially grateful for all her support and patience despite the incredibly difficult times during the pandemic. I would also like to thank my committee members: Dr. Austin Johnson and Dr. Stephanie Moore. Thank you for helping me reach the finish line.

ABSTRACT OF THE DISSERTATION

A Meta-Analysis of Single-Case Research on Mathematics Interventions for Students with Mathematics Learning Disabilities and Mathematics Difficulties

by

Luisana Suchilt

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, September 2022
Dr. Asha K. Jitendra, Chairperson

A meta-analysis of single-case design research was conducted to evaluate the effectiveness of mathematics interventions for elementary grade (K-5) students with mathematics difficulties (MD) and mathematics learning disabilities (LD). The study examined whether intervention effectiveness varied as a function of participant characteristics (i.e., grade level, MD/LD status), intervention characteristics (i.e., intervention type, intervention implementer, intervention duration, intervention group size, mathematics domain), and study quality (i.e., evidence of a functional relation). Results included 20 studies published from 2009 to 2021, with a total of 31 effect size estimates. The overall mean effect size estimate was 0.91 ($p < .01$), indicating significant positive outcomes for students with MD and LD after receiving mathematics interventions. Moderator analyses indicated that grade level and intervention type moderated treatment effects, such that students in Grade 3 showed higher treatment effects than students in Grade 2, and interventions that used explicit instruction yielded the highest effect sizes compared to interventions that used schema-based instruction, cognitive strategy instruction, and concrete representational abstract (CRA) instruction. Disability status, intervention duration, intervention group size, mathematics domain, intervention implementer, and study quality did not moderate intervention effects. However, findings from the moderator analyses should be interpreted with caution given the small number of studies within various categories. The present study provides further evidence regarding the effectiveness of mathematics interventions for students with MD and LD when using single-case research design. Limitations, directions for future research, and implications for practice are discussed.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Mathematics is of central importance in all aspects of life. Daily tasks such as managing time, budgeting money, cooking, and driving all involve mathematical skills. Likewise, jobs in various fields such as accounting, logistics, retail, and computer science require a level of mathematics competency. Moreover, mathematical skills serve as a foundation for students' academic success. Students who grasp basic arithmetic ideas and know how to apply them early in their education are better prepared to succeed as they progress through the grades (Bailey et al., 2014; Jordan et al., 2009). However, mathematics is challenging for most students in the United States.

Despite the high standards and expectations set forth by the Common Core Standards for Mathematics (CCSS-M), many students are not meeting the benchmarks for mathematical proficiency. According to the 2019 National Assessment of Educational Progress (NAEP), only 41% of fourth-graders and 34% of eighth-graders performed at or above proficient levels in mathematics, with percentages dropping to 24% for students in twelfth grade (National Center for Education Statistics, 2019). These percentages are lower for students with disabilities. For example, in 2019, only 17% of fourth-grade students with disabilities scored at or above proficient, compared to 45% of students without disabilities. In the eighth and twelfth grades, only 9% and 7% of students with disabilities scored at or above proficient in mathematics, compared to 38% and 27% for students without disabilities, respectively. Furthermore, compared to their typically-developing peers, students with disabilities show delays in the development of their mathematical abilities. For example, students with disabilities demonstrate difficulties with basic mathematics facts and calculations (Bouck & Park, 2018; Methe et al., 2012), word problem-solving (Alghamdi et al., 2020; Powell, 2011; Shin et al., 2020; Xin & Jitendra, 1999; Zhang & Xin, 2012), understanding and solving fractions (Fuchs et al., 2017; Jordan et al., 2017; Misquitta, 2011; Shin & Bryant, 2015), and algebra (Satsangi et al., 2016; Watt et al., 2016). These achievement gaps place students with disabilities at a further disadvantage compared to their typically achieving peers.

Mathematics difficulties can be persistent and continue throughout a student's schooling and into adulthood. For example, low mathematics achievement in kindergarten has been linked to slower growth rates and continuous struggles in mathematics throughout the elementary years (Jordan et al., 2009; Morgan, 2009). Mathematics difficulties also continue into the secondary grades. For instance, Bailey et al. (2014) found that whole number magnitude knowledge and whole number arithmetic in first grade predicted mathematics performance in middle school. Moreover, post high school, students with mathematics difficulties demonstrate lower college readiness (Lee, 2012), higher rates of unemployment, and lower income wages (Dougherty, 2003; Parsons & Bynner, 1997; Rivera, 1992). The concerning implications that low mathematics performance has throughout a student's educational trajectory and into adulthood calls for the identification and implementation of evidence-based interventions (EBI) to help increase mathematics achievement in students.

To address the needs of students struggling with mathematics, there have been calls in the early 2000s to use scientifically based research practices (No Child Left

Behind Act of 2001, 2002). As a result, there has been an increase in intervention research for improving the learning of students with mathematical learning disabilities (LD) using evidence-based interventions (EBI) and best practices (Jayanthi et al., 2008; Gersten et al., 2009; Fuchs et al., 2021). For example, in 2009, the Institute of Education Sciences (IES) released a practice guide to highlight and describe eight recommendations to assist educators in addressing the needs of students with mathematics difficulties. These recommendations were evaluated following the What Works Clearinghouse (WWC) guidelines, which are used to determine whether interventions are supported as effective by rigorous research. Recently, the IES released updated recommendations that focused primarily on effective small-group interventions for students in K-6 (Fuchs et al., 2021). Additional practice guides have focused on developing effective fraction instruction for kindergarten through 8th grade (Siegler et al., 2010), improving mathematical problem-solving in grades 4 through 8 (Woodward et al., 2012, 2018), and teaching strategies for improving algebra knowledge in middle and high school students (Star et al., 2015, 2019).

Despite an increase in intervention research, mathematics achievement gaps have not decreased (Dossey et al., 2016; Rittle-Johnson & Jordan, 2016). In response to the low achievement of students in mathematics, many schools have begun to deliver interventions through a multi-tiered system of supports (MTSS) framework (Fuchs & Vaugh, 2012). MTSS is an instructional framework that allows for the implementation of interventions based on student need. Within an MTSS framework, instruction is generally delivered across three tiers. Tier 1 involves high-quality core instruction that all students receive. Students who do not respond to core instruction are offered instructional support in Tier 2, which encompasses supplemental instruction in the form of targeted, small-group instruction. Tier 3 offers more intensive and individualized instruction for students who do not respond to Tier 2 instruction. MTSS is grounded on the implementation of EBIs and best practices. Because of this, examining the effectiveness of instructional interventions and practices is important for supporting student learning. As a result, the goal of this dissertation is to conduct a meta-analysis to evaluate the effectiveness of mathematics interventions for students struggling in mathematics.

Prior meta-analyses have examined the effects of mathematics interventions for students with mathematics difficulties (MD) and mathematics learning disabilities (LD) (e.g., Chodura et al., 2015; Dennis et al., 2016, Stevens et al., 2018; Gersten et al., 2009; Jitendra et al., 2020; Kroesbergen & Van Luit, 2003). These studies have generally found that students show improvement in mathematics outcomes when receiving intervention supports. However, prior meta-analyses of the effects of mathematics interventions have primarily focused on evaluating group-design studies. It is equally important to synthesize the effectiveness of mathematics interventions using single-case design (SCD) research, because the methodology is well suited to investigate single cases with respect to understanding how each student with MD or LD respond to the intervention (Kazdin, 2011).

Single-case design methodology is recognized as a rigorous method for testing the effectiveness of educational practices (Horner, 2005; WWC, 2020). A key feature of SCD is the systematic and continuous measurement of the outcome variable, allowing for multiple demonstrations of the effect as the intervention is systematically introduced and withdrawn. Similar to randomized control group designs, SCD research allows for experimental control, which is critical when establishing evidence-based practices. Furthermore, as each participant in SCD research serves as his or her own control, we can examine the effectiveness of interventions at the individual level. This may be especially important when testing interventions for students with heterogenous needs, such as students in special education or students who are English Language Learners (Horner et al., 2005; Kazdin, 2011). Lastly, SCD research may also be more cost effective for identifying and evaluating interventions that may later be studied in a large scale through group design (Kazdin, 2011; Baker at al., 2018).

The primary aim of this dissertation is to conduct a meta-analysis of single-case design studies to evaluate the effectiveness of mathematics interventions for elementary grade students with MD and LD. The proposed study will also examine whether intervention effectiveness varies as a function of participant characteristics (i.e., grade level, MD/LD status), intervention characteristics (i.e., intervention type, intervention implementer, intervention duration, group size, mathematics domain), and study quality (i.e., evidence of a functional relation). The current study will extend the literature on mathematics interventions by informing mathematics instruction for students with MD and LD.

Literature Review

In this chapter, I first describe characteristics of students with mathematics learning disabilities (LD) and mathematics difficulties (MD), as well as consider previous research on identifying students as having MD and LD. Second, I review previous group design meta-analytic studies examining the effectiveness of mathematics interventions for students with LD and MD. Lastly, I review relevant prior SCD meta-analyses of mathematics intervention for students with LD and MD.

Characteristics of Students with LD and MD

Students who consistently demonstrate low achievement in mathematics generally fall within one of two categories. These students may be classified as (1) having a learning disability in mathematics (LD) or (2) demonstrating mathematics difficulties (MD). Students with LD may have a diagnosis of a specific learning disability in mathematics and also have an Individualized Education Plan (IEP) with specific goals in mathematics as part of special education service provisions. In contrast, students with MD do not have an official diagnosis. Instead, the term MD represents a broader construct than LD in mathematics and is used to refer to students with persistent low mathematics performance. Students with MD are considered to be at-risk of having future mathematics difficulties and developing a disability in mathematics. However, the criteria used to define and identify students with mathematics challenges varies across the literature.

Definition of LD and MD

A lack of consensus exists in defining and characterizing students with mathematics learning disabilities (LD). Although the operational definition of LD has been inconsistent across the literature, the two most common criteria for defining LD are (1) a discrepancy between IQ and performance level and (2) performance at or below a cutoff criterion on a mathematics measure. In research studies, however, cutoff criteria vary, often ranging from the 10^{th} percentile to the 35^{th} percentiles (Geary, 2011; Price & Ansari, 2013). Unfortunately, variations in cutoff scores pose issues when identifying and determining whether an individual has a learning disability in mathematics (LD) or has mathematics difficulties (MD). For example, Murphy et al. (2007) categorized all students performing below the 25^{th} percentile as having LD, whereas Geary (2011) only categorized students as LD if they performed below the 10^{th} percentile, while performance between the 11^{th} and 25^{th} percentile was used to classify students as low achieving or as having mathematics difficulties (MD). In sum, there is not always a clear distinction between students with LD and MD.

Discrepancies also exist when identifying and defining students who do not have a mathematics learning disability but are demonstrating persistent mathematics difficulties (MD). The literature often refers to these individuals as students who are at risk of developing a learning disability in mathematics or students with MD (Baker et al., 2002; Dennis et al., 2016; Mazzoco & Thompson, 2005). Criteria for identifying these students also vary across studies, but generally involves meeting one or both of the following: (1) score below average on tests of mathematics achievement, and (2) demonstrate low performance in mathematics based on teacher recommendation (Baker et al., 2002; Hinton & Flores, 2019; Gersten et al., 2005; Sharp & Dennis, 2017; Xin & Zhang, 2009). Again, issues arise when using cutoff scores to categorize a student as having MD. For example, Gersten et al. (2005) opted not to use the term *mathematics disabilities* and instead used the term *mathematics difficulties* to encompass all students performing below the 35th percentile on a mathematics measure. Therefore, in this study, students with a learning disability in mathematics were included with students who did not have a disability but were experiencing mathematics difficulties. The inconsistent criteria and overlap in cutoff scores across studies makes it challenging to separate these students into two distinct groups.

Many researchers have used the term MD to refer to both students with a mathematics learning disability and students with persistent low mathematics performance (Gersten et al., 2005; Powell et al., 2019). While students with LD and MD represent individuals with heterogenous mathematics difficulties, they also represent students with mathematics skills existing on a continuum, as opposed to students with entirely different skills (Andersson, 2008; Geary et al., 2012; Kroesbergen & Van Luit, 2003). As a result, researchers often study both groups of students as a whole, but also perform separate analysis for each group (Andersson, 2008; Dennis et al., 2016; Lein et al., 2020; Xin & Jitendra, 1999). Recognizing that students with LD and MD are not homogeneous is important, however recognizing that both groups of students benefit from mathematics instruction and interventions that will help the remediation of specific mathematical skills is just as meaningful (Geary et al., 2012; Gesten et al., 2005). In the following section, I describe domain-general and domain-specific skills that contribute to these students' low achievement in mathematics.

Domain-General Skills

Studies have consistently demonstrated that students with LD and MD are characterized by impairments in cognitive skills related to executive functioning, specifically working memory and attention processes. Deficits in working memory have repeatedly been linked to students who struggle with mathematics (Andersson, 2010; Geary, 2011; Geary et al., 2007; Passolunghi & Siegel, 2004), and there is evidence that mathematics performance is also affected by attention processes (Geary, 2011; Hassinger-Das et al., 2014; Peng et al., 2015; Swanson et al., 2015; Swanson & Sachse-Lee, 2001). Compared to typically developing children, children with mathematics difficulties show impairments in working memory tasks involving both verbal or numerical information (Passolunghi & Siegel, 2004; Swanson et al., 2015). Children with MD are more likely to perform lower on a digit span backward task, which requires active processing of numerical information, and also make more intrusion errors (Geary et al., 1999; Passolunghi & Siegel, 2004; Shin & Bryant, 2013). In other words, when recalling verbal or numerical information, children with MD struggle with inhibiting irrelevant information; such difficulties have been attributed to deficits in the central executive, a component of working memory responsible for regulating information being received (Passolunghi & Siegel, 2004; Peng et al., 2015).

Executive functioning skills are closely linked to mathematics skills. Hassinger-Das et al. (2014) examined how attention problems and executive functioning, including inhibition and working memory, were related to first grade students' number sense abilities. Number knowledge in kindergarten predicted number abilities in the first grade, and this relation was partially explained by students' attention problems and executive functioning, such that students with low number sense in kindergarten who demonstrated lower executive functioning and more attention problems performed worse in first grade mathematics. One aspect of executive processing is being able to access information from long term memory to retrieve basic arithmetic facts necessary when solving mathematical problems. Children with LD are more likely to commit errors when retrieving information from long term memory and because of this they are more likely to commit counting errors and use ineffective counting procedures (Geary, 2004; Geary, 2011; Shin & Bryant, 2015; Swanson & Lee, 2001). For example, when adding 3 + 5, a struggling student is more likely to count starting from one, instead of counting starting from the larger addend. Children who show low achievement in mathematics also demonstrate lower speed of number processing compared to their typically developing peers, this slow processing is linked to attention processes of the central executive. However, more research is needed to understand the differences in the mechanisms explaining the mathematics challenges of students with MD and students with LD (Geary, 2011).

Domain-Specific Mathematics Skills

Compared to typically achieving children, students with LD and MD generally demonstrate lower performance across mathematics tasks (Passolunghi & Siegel, 2004). They commit more procedural errors when solving simple and complex arithmetic problems and take longer to complete a task (Anderson, 2008; Burns et al., 2010; Geary, 2011; Passolunghi & Siegel, 2004). Deficits in working memory contribute to ineffective counting strategies, thus delaying the mastery of basic mathematical skills (Geary, 2004). Students with LD and MD are also more likely to use their fingers when counting, whereas typically achieving children are more likely to count silently in their head by the time they are in first grade (Geary, 2011). Furthermore, children with MD are slower and less accurate in identifying the correct mathematical operations (e.g., addition, subtraction) required to solve a problem and have trouble switching between operations. For example, they may add when they are supposed to subtract (Jordan and Montani, 1997 cited in Geary 2011). Difficulties remembering and retrieving basic arithmetic facts (Geary, 2011; Passolunghi & Siegel, 2004) and making calculation errors when solving complex problems are also common in students with LD and MD (Geary, 2011; Mazzocco & Thompson, 2005). These students are more likely to struggle with representing the magnitude of numbers and understanding the concept of number lines (Geary et al., 2008).

Children with LD and MD show delays in the development of their mathematical abilities. They consistently demonstrate impairments in conceptual and procedural knowledge (Anderson, 2008; Burns et al., 2010; Geary, 2004). Difficulties with number sense and understanding the concept of number lines (Geary, 2008; Hassinger-Das, 2014), counting strategies (Andersson, 2008; Mazzocco & Thompson, 2005), basic mathematics facts and calculations (Bouck & Park, 2018; Methe et al., 2012), word problem-solving (Alghamdi et al., 2020; Powell, 2011; Shin et al., 2020; Xin & Jitendra, 1999; Zhang & Xin, 2012), understanding and solving fractions (Jordan et al., 2017;

Fuchs et al., 2017; Misquitta, 2011; Shin & Bryant, 2015), and algebra (Satsangi et al., 2016; Watt et al., 2016), place these students at a disadvantage compared to their typically achieving peers. As a result, addressing and studying effective instructional practices and interventions to meet the needs of these students is essential for improving their mathematics outcomes (Dennis et al., 2016; Gersten et al., 2009; Kroesbergen & Van Luit, 2013; Nelson & Powell, 2018).

Prior Meta Analyses of Mathematics Interventions

Given the well-documented inadequate mathematics performance of students with LD and MD, several prior meta-analyses have examined the effects of mathematics interventions for these students. These meta-analyses have evaluated both group design (e.g., Chodura et al., 2015; Dennis et al., 2016, Stevens et al., 2018; Gersten et al., 2009; Jitendra et al., 2020; Kroesbergen & Van Luit, 2003) and single case design (SCD) studies (e.g., Lei et al., 2020; Shin et al., 2020). Meta-analyses involving group design and SCD research have focused on specific mathematical content instruction such as early numeracy (Nelson & McMaster, 2018), addition and subtraction calculation skills (Burns et al., 2002; Methe et al., 2012), word problem solving (Shin et al., 2020; Xin & Jitendra, 1999; Zhang & Xin, 2012; Zheng et al., 2013), fractions (Ennis & Losinski, 2020), and algebra (Hughes et al., 2014). Other meta-analyses have focused on the effects of mathematics interventions involving the use of manipulatives (Peltier et al., 2020) and mathematics interventions for secondary students with MD (Jitendra et al., 2018).

These studies have revealed positive effects of mathematics interventions on certain outcomes for school-aged children. Results suggested that additional factors moderate the effect of mathematics interventions on students' mathematics achievement. Such factors have included grade level, disability status, intervention implementer, intervention duration, and instructional practices or approaches. In the following sections, I review the meta-analytic studies that examined the effectiveness of mathematics interventions for students with LD and MD, detailing variations in coding schemes and methodological decisions.

Group Design Meta-Analyses

In an early meta-analysis, Kroesbergen & Van Luit (2003) examined the effects of mathematics interventions for students with special needs (e.g., low achieving, learning disability) in grades K-6. A total of 34 group design studies and 24 SCD studies were categorized in terms of: (1) instructional approach (i.e., direct instruction, self-instruction, and mediated or assisted instruction) and (2) intervention medium (i.e., teacher directed or computer-assisted interventions), and (3) mathematics domains (i.e., preparatory arithmetic, basic facts, and problem solving). Results indicated that students with learning disabilities showed the largest treatment effect, followed by low achieving students, and students with other disabilities (e.g., behavior or attention disorder). In regard to instructional practices, students benefited the most from self-instruction (d = 1.45), followed by direct instruction (d = 0.91), and mediated instruction (d = 0.34). Furthermore, interventions implemented by a teacher (d = 1.05), resulted in larger treatment effect compared to computer-delivered interventions (d = 0.51). Although interventions targeting basic facts (d = 1.14) were the most commonly implemented and demonstrated the highest effect sizes, there were no significant differences in effect sizes across the different mathematics domains. Lastly, duration of the intervention significantly moderated the intervention effects, such that longer interventions had less effect on student outcomes, compared to shorter interventions ($\gamma = -0.01$). Several limitations of this study include: (1) excluding unpublished studies, (2) not disaggregating results for group design and SCD studies, given that SCD studies tend to produce larger effect sizes, (3) averaging effect sizes within a single study without taking into account the issue of dependency, and (4) including studies of complex interventions that involved more than mathematics instruction (e.g., parent intervention, rewards) without addressing how this may bias effect sizes.

Gersten et al. (2009) examined the literature on mathematics intervention studies for school-aged students with an identified learning disability. The authors were specifically interested in assessing the extent to which instructional practices contributed to positive mathematics outcomes in students with learning disabilities. Their study included 42 group-design intervention studies (i.e., randomized control trials and quasi-experimental design). Studies were coded for four instructional categories: (1) instructional approach or curriculum design (e.g., explicit instruction, use of heuristics, student verbalizations of mathematical reasoning), (2) providing formative assessment data and feedback to teachers on students' mathematics performance (e.g., student progress monitoring data, instructional recommendations) (3) providing formative assessment data and feedback to students (e.g., feedback on performance), and (4) peer-assisted mathematics instruction. The largest effects were found for interventions that included explicit instruction ($g = 1.22$), use of heuristics ($g = 1.56$), student verbalizations of their mathematical reasoning (g = 1.04), using visual representations while solving problems (g = 0.41), sequencing and or providing a range of examples (g = 0.82), providing ongoing data and feedback to teacher on student's performance (g = 0.23), and providing students with information on their progress (g = 0.23). However, one limitation of this meta-analysis was that given the complexity of interventions, studies were coded for more than one instructional component, making it difficult to isolate each instructional component's unique contribution.

Chodura et al. (2015) conducted a meta-analysis of 35 group design studies to determine the extent to which elementary school students benefited from mathematics interventions, and to investigate the characteristics that make an intervention effective. The authors were primarily interested in three subgroups of students: students with learning disabilities (i.e., students with mathematics difficulty and a comorbid disability), students with dyscalculia (i.e., students with an identified mathematics disability based on ICD or DSM criteria) and students at risk of developing dyscalculia (i.e., students performing below the 26th percentile on a standardized mathematics test). The authors also examined the role of intervention components, including the type of instruction (i.e., direct, strategy, assisted, self) and the mathematics domains targeted (e.g., basic arithmetical competencies, problem solving). Results indicated an overall positive estimated effect size (g = 0.83). However, subgroup analysis indicated that only students who were at risk of developing dyscalculia showed significant treatment effects (g = 0.80). These students benefited from interventions that incorporated direct instruction (g = 0.60), promoted problem solving (g = 0.67), were computer-based (g = 0.80), were delivered individually (g = 0.91), and were not adaptive (g = 0.61). One limitation of this study was the overlap among the three student subgroups, as it was not clear how to differentiate between students who had a learning disability and students who had dyscalculia.

Dennis et al. (2016) conducted a meta-analysis of mathematics intervention studies for school-aged children with mathematic difficulties (i.e., students who performed below the 25^{th} percentile on a mathematics screening test). The authors examined the role of participant characteristics and intervention parameters in predicting mathematics outcomes. A total of 25 group design studies were included in their analysis. The estimated overall effect size indicated positive mathematics outcomes for students with MD (g = 0.53). Additional analyses were conducted to examine the potential impact of moderators. Influential participant characteristics included grade level and mathematics skill level, such that students at the elementary level (g = 0.57) benefited more from the interventions than students in kindergarten (g = 0.30). None of the studies included interventions for students at the secondary level. Additionally, students who fell at or below the 35^{th} percentile when identified with MD (g = 0.70), demonstrated more improvements than students who were above the 35^{th} percentile when identified (g = 0.14). That is, students with a larger degree of mathematics difficulty benefited more from intervention. Peer-assisted learning (g = 0.82) and explicit teacher-led instruction (g = 0.76) were the most effective instructional approaches, followed by providing student data to teachers (g = 0.63) and using technology (g = 0.39). Interventions were more effective when delivered by a researcher (g = 0.81) and when they were implemented in a large (g = 0.83) or in a small group (g = 0.70). Moderator analyses indicated that task difficulty (β = 1.72), elaboration (e.g., explicit instruction, modeling; (β = 1.27), and small group intervention delivery (β = 1.30) were significantly predictive of treatment effects.

Stevens et al. (2018) conducted a meta-analysis to evaluate the effects of 25 mathematics interventions for students in Grades 4-12 with mathematics difficulties. The term mathematics difficulties (MD) was used to refer to students with a specific learning disability in mathematics or students who demonstrated low mathematics achievement but did not a have a formal diagnosis. Teacher recommendation or performance below the 40^{th} percentile on a mathematics measure was used to determine whether a student was demonstrating low mathematics achievement. The authors also investigated the role of moderator variables in determining intervention effectiveness. These variables included: intervention characteristics, student characteristics, mathematics content, and study quality. Results indicated great variability across the studies (g = -0.66-4.65), with an estimated non-significant pooled effect of 0.49 after excluding studies with a small sample size. Furthermore, after adjusting for small-study effects it was found that students showed stronger responses when receiving 15 or more hours of intervention (g = 0.79) and demonstrated higher treatments effects when receiving instruction in fractions (g = 0.74), as opposed to operations (g = -0.76). However, grade, number of intervention sessions, group size, and study quality did not predict effect size. One notable challenge from this meta-analysis was the large variability across estimated effect sizes, suggesting the need to explore additional moderator variables to better understand the unexplained heterogeneity across the studies.

In a recent meta-analysis, Jitendra et al. (2020) examined the effectiveness of Tier 2 mathematics interventions for students with mathematics difficulties. They explored the effects of: (1) MD status (i.e., screening based on a single or multiple measure), (2) grade level, (3) intervention model type (e.g., strategy instruction, problem-structure instruction), (4) interventionist, (5) group size, (6) duration of intervention, (7) quality of study, and (8) outcome measure. A total of 39 group design studies of students in pre-kindergarten through Grade 12 were included in the meta-analysis; however, a large proportion of students in the studies were elementary-aged (91%). Similar to Stevens et al. (2018), students were determined to have a mathematics difficulty if they (a) scored at or below the 40th percentile on a screening mathematics test or (b) were eligible for Tier 2 services. Eligibility for Tier 2 services was based on specific criteria set forth by the authors of each study. Results indicated an overall moderate significant estimated effect size ($g = 0.41$). Intervention model type ($g = 0.42$), group size ($g = 0.29$), and type of outcome measure ($g = 0.26$) were significant predictors of student outcomes. Specifically, problem structure instruction only and intervention groups of two or three students were associated with positive significant treatment effects. In contrast, standardized measures yielded a significant negative effect size. Further, MD status, grade level, interventionist, duration of intervention, and quality of study did not significantly impact the effects of Tier 2 interventions. This study's limitations include the exclusion of unpublished studies and the need to consider additional moderator variables, such as the mathematics domains targeted by the interventions.

Review of Relevant Prior SCD Meta-Analyses

Meta-analyses examining the effectiveness of mathematics interventions for students with learning disabilities and mathematics difficulties in the elementary grades have generally found positive mathematics outcomes for students. However, meta-analyses that have synthesized this literature have primarily focused on group-design studies. As such, the following section will review the scant research on meta-analyses that have focused solely on mathematics intervention studies that employed a SCD research design. Specifically, I will (1) identify the mathematics domains that have been studied, (2) address the distinct criteria used to determine MD and LD eligibility across studies (3) examine the intervention characteristics that have been studied (4) determine the extent to which quality of study was assessed and (5) discuss the different use of effect sizes. Lastly, variations and inconsistencies across the studies will be addressed and discussed.

Burns et al. (2010) conducted a meta-analysis of 17 single-case studies to assess the effectiveness of acquisition and fluency interventions for students in Grades 2-6 with mathematics skills in the frustration and instructional levels. Acquisition interventions involved teaching and modeling concepts and were meant to increase accuracy. In contrast, fluency interventions targeted students who had acquired the skill but required practice to reach proficiency. The authors sought to determine whether a skill-by-treatment effect was present, such that students in the frustration level would benefit more from acquisition interventions (e.g., modeling, guided practice), while students in the instructional level would benefit more from fluency-based interventions (e.g., practice, goal setting, feedback). Baseline data from each study was used to identify students as being in the instructional or frustration level. Levels were determined based on students' digits correct per minute (dcpm). Results indicated a large treatment effect of acquisition interventions on students with skills in the frustration level (phi = 0.84, CI = 0.76-0.93) and a moderate effect on students in the instructional level (phi = 0.49, CI = 0.29-0.70). In contrast, results suggested a small to moderate effect of fluency interventions on students within the frustration level (phi = 0.47, CI = 0.25 – 0.68; PAND = 75%). The effect of fluency interventions on students in the instructional level could not be interpreted due to a small number of studies. Findings support that intervention level should match level of skill proficiency.

Methe et al. (2012) conducted a meta-analysis of 11 single-case studies that targeted the addition and subtraction skills of students in grades K-6. In order to be included in the meta-analysis, the studies must have included "participants with achievement problems or high-incidence disabilities (e.g., below grade expectations for mathematics performance, specific learning disabilities)" (p. 235). The authors also examined intervention type and degree of experimental control as potential moderators. Experimental control was determined through visual analysis and was quantified on a continuous 10-point scale. In order to earn all 10 points, the study needed to show prediction, affirmation of the consequent, verification, and replication. Findings indicated overall positive effect sizes (IRD = 0.59 and PAND = 0.66). Mathematics interventions were more effective for younger students. Specifically, students in the second (IRD = 0.88), third (IRD = 1.00), and fourth grade (IRD = 0.80) showed more improvement, compared to students in the fifth (IRD = 0.43), and sixth grade (IRD = 0.58). Fluency-based interventions were the most effective (IRD = 0.97), followed by accuracy-based interventions (IRD = 0.94), and interventions with combined components (IRD = 0.65). However, it should be noted that only three studies assessed fluency-based interventions and only one assessed accuracy-based interventions. Lastly, the authors found a negative relationship between experimental control and treatment effects.

Peltier et al. (2020) narrowed their focus and only examined the effect of interventions that used manipulatives to enhance mathematics achievement. Manipulatives were defined as concrete objects that could be used to support mathematical learning. A total of 53 single-case studies were included in the meta-analysis. In order to be included, studies had to target students in Pre-Kindergarten-Grade12 who were at risk or identified with a disability. The authors also assessed each study for methodological quality using a coding sheet that aligned with the WWC Pilot Single-Case Design Standards (Kratochwill et al., 2010). Overall, findings indicated that the use of manipulatives can help improve mathematical outcomes in students at risk and students identified with a learning disability (Tau-*U* = 0.34-1.00; BC-SMD = 0.03-18.58). In terms of participant characteristics, age did not moderate intervention effects. Moreover, interventions were less effective for students with emotional and behavioral disorders (EBD), however, the authors suggest caution when interpreting this finding given the small sample size of students with EBD. Shifting to the analysis of intervention components, results indicated that the format of manipulative (i.e., concrete vs. virtual), type of manipulative (i.e., bland vs. rich), interventionist (i.e., teacher vs. researcher), and mathematical concepts (e.g., computation, algebra, problem-solving) did not moderate the treatment effect. Lastly, study quality and study design did not moderate intervention effects.

In a recent synthesis, Shin et al. (2020) conducted a meta-analysis of 20 single-case studies to examine the effectiveness of word problem solving interventions on students with learning disabilities in Grades 1-12. Results from this study indicated that students with learning disabilities benefited from word problem-solving instruction (BC-SMD = 4.52). The authors also examined moderators. First, the methodological quality of all studies was assessed using the quality indicators put forth by the Council for Exceptional Children (CEC) 2014. While study quality was positively associated with the weighted effect size estimate, the relationship was not significant. Next, the study examined whether the extent to which studies met the Common Core State Standards for Mathematics (CCSSM) and whether meeting a specific standard moderated treatment effects. Results indicated that interventions that targeted Operations and Algebraic Thinking (OA; BC-SMD = 2.98) and Number System (BC-SMD = 3.04) led to significant improvements in word-problem solving. The largest treatment effects were for interventions that targeted geometry (BC-SMD = 11.80) and Expressions and Equations (BC-SMD = 6.21); however, these improvements were not statistically significant. Lastly, the authors investigated whether implementing CCSS mathematical standards had an impact on students' word-problem solving outcomes. Studies that incorporated the following standards showed significant improvements in word problem solving: (1) make sense of problems and persevere in solving them (BC-SMD = 3.98), (2) reason abstractly and quantitively (BC-SMD = 4.52), (3) construct viable arguments and critique the reasoning of others (BC-SMD = 4.24), (4) model with mathematics (BC-SMD = 4.52), (5) use appropriate tools strategically (BC-SMD = 4.52), and (6) attend to precision (BC-SMD = 3.98). The largest effect was found for studies that involved looking for and making use of the structure of the problem. However, this finding was not significant.

Lei et al. (2020) also synthesized the literature on mathematics word problem solving (WPS) interventions but focused specifically on interventions for English Language Learners (ELL) in Grades K-12 with mathematics difficulties (MD) and learning difficulties (LD). A total of 10 single case design intervention studies were included in the study. The authors examined both participant characteristics and intervention characteristics. Given the critical role of reading ability in solving word-problems, the authors also examined the extent to which students benefited from interventions that incorporated reading comprehension instruction. The authors examined the following intervention characteristics: instructional focus (i.e., WPS only, WPS and reading comprehension, or reading comprehension only) and intervention features (i.e., intervention duration, agent, and setting). Quality of study was also evaluated using the What Works Clearinghouse (WWC) Design Standards (Kratochwill et al., 2013). Results indicated an overall moderate effect size (Tau-U = 0.81). Students with a learning disability (Tau-U = 1.00) showed a greater treatment effect than students with mathematics learning difficulties (Tau-U = 0.78). Interventions that targeted mathematics instruction only were more effective (Tau-U = 1.00) than instruction in both reading comprehension and mathematics (Tau-U = 0.81), and this was followed by instruction in reading comprehension only (Tau-U = 0.71) (e.g., teaching mathematics vocabulary, paraphrasing word problems). When comparing the effect of grade, it was found that students in fourth grade (Tau-U = 1.00) showed larger improvement than students in third grade (Tau-U = 0.75). Furthermore, the authors examined the influence of specific intervention characteristics. Interventions implemented by teachers and interventions with a total duration of fewer than 10 hours showed the largest effect size. No difference was found between interventions implemented in a one-on-one and small group setting.

Mathematics Domains/Outcomes

The abovementioned meta-analyses that have been conducted to synthesize the evidence base of SCD research on mathematics interventions have been narrow in focus. Each meta-analysis has centered on a specific mathematics domain or ability (e.g., word problem-solving). First, Burns et al. (2010) focused on computation and fluency skills. The authors only examined addition, subtraction, multiplication, and division interventions in which digits correct per minute (dcpm) could be calculated. Digits correct per minute was used as a measure of computation accuracy and fluency skills. In other words, the study targeted acquisition and fluency interventions, which aimed to improve accuracy (in computation) and proficiency (fluency). Likewise, Methe et al. (2012) also focused on mathematics computation but only targeted interventions in the areas of addition and subtraction, as they considered these skills to be distinct from multiplication and division skills. Similar to Burns et al. (2010), the authors were also interested in examining the effect of accuracy and fluency-based interventions on computation skills.

On the other hand, Lei et al. (2020) and Shin et al. (2020) investigated the effects of word-problem solving instruction on students' word problem solving skills. Lei et al. narrowed their focus to students who were ELLs, and as a result they also examined the extent to which instruction in reading comprehension contributed to mathematics outcomes. Lastly, Peltier et al. (2020) did not focus on a specific mathematics domain but instead sought to examine the effect of using manipulatives to improve mathematics outcomes. As such, the study included interventions that addressed various mathematics domains, including, algebra, basic facts, computation, early numeracy, fractions, geometry, money, and problem-solving.

Findings from Burns et al. and Methe et al. both indicated that students' computation and fluency skills improved as a result of intervention. However, Burns et al. also found that students in the frustration level benefited more from accuracy-based interventions than fluency-based interventions. Likewise, Shin et al. and Lei et al. found positive outcomes on students' word-problem solving. In addition, Lei et al. found that students showed more WPS improvement when instruction focused only on mathematics, as opposed to a combination of reading comprehension and mathematics, or reading comprehension only. Lastly, results from Peltier et al. indicated that the use of manipulatives led to improvements across mathematics domains.

Participant Characteristics

Prior meta-analyses have investigated the effects of mathematics interventions for students with mathematics learning disabilities and mathematics difficulties. However, variations exist in defining LD and MD. Likewise, studies have reported inconsistent mathematics outcomes for these students. The next section will highlight the different criteria used to determine whether a student met eligibility for LD and MD, as well as the variations in findings.

LD/MD Status. Burns et al. (2010) examined intervention studies for students whose mathematics skills were in the frustration and instructional levels. Their sample included students with a learning disability in mathematics (10.9%), intellectual disability (21.8%), behavioral disorders (12.7%), or were not identified with a disability (54.5%). The authors indicated that 75 percent of the studies used assessment data to identify students and determine an appropriate intervention. These studies used either a screening measure, a standardized mathematics test, a placement test, or a combination of data to identify students. The authors did not conduct analyses for each disability category. Instead, students were categorized into two levels: frustrational and instructional. These levels were determined based on students' median digits correct per minute score.

Methe (2012) targeted students with "achievement problems and high incidence disabilities" (p. 235): that is, students with low performance in mathematics and students with specific learning disabilities. Their sample consisted of 47 students, of which 41 were referred for mathematics performance below grade or below age level. It was not clear as to what difficulties comprised the rest of the sample. Further, the authors did not specify what criteria were used to determine whether a student had an achievement problem in mathematics (e.g., screening measure, standardized test) or how students were identified as having a learning disability

Peltier et al. (2020) examined interventions for students at risk or identified with a learning disability. Their sample included students in the following categories: at risk, autism spectrum disorder, developmental delay, emotional and behavioral disorder, intellectual disability, other health impairment, and specific learning disability. The authors did not specify the criteria used to determine whether a student was at risk or had a learning disability. However, they reasoned that students at risk were included in their study because they showed similar performance levels as students with learning disabilities on universal screener data. Likewise, they note that within a multi-tiered system of supports framework, students who are at risk would be receiving similar services along with students with learning disabilities. Further, unlike Burns et al. and Methe et al. this study examined disability status as a potential moderator and found that mathematics interventions were less effective for students with behavioral disorders. This finding highlights the need to consider disability status as a potential moderator.

In Shin et al. (2020), participants were students who had been identified as having a learning disability by their school district. The authors report that all students (n = 82) had difficulties in mathematical word problem-solving, based on their screening test scores. However, 16 students had additional difficulties in reading or writing. The authors acknowledge that the studies used different criteria in determining whether a student had a learning disability. For example, some studies used the IQ-achievement discrepancy, while others determined eligibility based on performance on a standardized mathematics test, or through the Response to Intervention (RTI) process.

Lei et al. (2020) evaluated interventions for ELL students with learning disabilities and mathematics difficulties. The authors stated that to be included in the meta-analysis, the study must include at least one participant who was an English Learner identified with LD or MD. In total, 87% of participants had MD and 13% were diagnosed with LD. The term MD was used to refer to students with learning difficulties in mathematics and students with low mathematics achievement. It was not clear as to how students were determined to have mathematics difficulties or a learning disability.

Altogether, findings from these meta-analyses suggest that students with mathematics difficulties benefit from intervention. Burns et al. and Methe et al. both found that students with low mathematics performance and learning disabilities showed improvements in computation skills after receiving intervention instruction. Results from Shin et al. and Lei et al. indicated that students with learning disabilities benefited from word problem-solving instruction. However, Lei et al. specifically found that students with a learning disability showed a greater treatment effect than students with mathematics difficulties. Lastly, Peltier et al. found that interventions that used manipulatives were effective for students with mathematics difficulties and students with learning disabilities. However, they were less effective for students with emotional and behavioral disorders. The authors suggest caution when interpreting this finding given the small sample size of students with EBD.

Grade. Burns et al. (2010) found that acquisition and fluency interventions targeting computation skills were effective for students in Grades 2-6 with mathematics skills in the frustration and instructional levels. No significant differences were detected across grades. However, Methe et al. found differences, such that basic computation interventions were more effective for younger students. Specifically, students in grades 2-4 showed more growth than students in grades 5-6. Shin et al. and Lei et al. both examined the effectiveness of word problem-solving interventions on students in Grades K-12. While Shin et al. did not report findings across grade level, results from Lei et al. showed that students in their sample (i.e., Grades 2-5) all benefitted from intervention. However, students in fourth grade showed significantly larger improvements than students in third grade. Lastly, Peltier et al. reported that participant grade level (i.e., preschool, elementary, intermediate, secondary) did not moderate intervention effects. Differences in findings across these studies could plausibly be explained by differences across the samples and differences in the intervention type (e.g., word-problem solving vs. manipulatives).

Intervention Components

Intervention Type and Components. Burns et al. evaluated a total of eight different interventions that were categorized into two types: acquisition or fluency intervention. These interventions included Great Leaps for Math, interspersal technique, self-monitoring, contingent reinforcement for increased performance, taped problems, timed warm-up probes, cover-copy-compare, and a combination of timings, peer-feedback, and positive practice overcorrection. The authors did not address specific instruction components or instructional practices.

Methe et al.'s meta-analysis included five intervention types: contingent reinforcement, cover-copy-compare, interspersal, speed-based intervention, combined intervention, and concrete-representational-abstract. The interventions were mostly focused on behavior (e.g., contingent reinforcement) or on improving fluency (e.g., cover-copy-compare), instead of being focused on teaching a mathematics skill. As a result, it is not clear as to the extent to which treatment effects were driven by changes in behavior, as opposed to changes in mathematics instruction, as the authors did not conduct separate analyses by intervention type. Likewise, many of the included studies assessed interventions that did not directly teach mathematics skills. For example, one intervention involved goal setting, and three others involved only cover-copy-compare (CCC), or a combination of CCC with another intervention. Therefore, there is no clear understanding of whether students' improvements were due to the remediation of specific mathematics skills, behavior modifications, or both. Further, the authors did not indicate specific instructional practices that comprised each intervention.

Shin et al. reported that the primary intervention types used across their studies to improve word problem-solving were: self-regulated, mnemonic strategy, model-drawing strategy, schema-instruction, explicit inquiry-based instruction, video modeling, and use of manipulatives. However, the authors did not examine differences across intervention types. Instead, they identified the extent to which the interventions met the Common Core State Standards for Mathematics (CCSSM) practice standards. Eight practice domains were coded: (1) make sense of problems and persevere in solving them, (2) reason abstractly and quantitively, (3) construct viable arguments and critique the reasoning of others, (4) model with mathematics, (5) use appropriate tools strategically, (6) attend to precision, (7) look for and make use of structure, and (8) look for and express regularity on repeated reasoning.

Lei et al., (2020) also evaluated word problem-solving intervention but given the focus on ELL students, the authors coded their interventions differently. Interventions were categorized by instructional focus: WPS only, WPS and reading comprehension, or reading comprehension. Findings indicated that ELL students benefited from instruction in both mathematics and reading; however, they benefited the most from instruction in mathematics only, and this effect was significantly different from instruction in reading comprehension only.

Peltier et al. (2020) were interested in interventions that primarily used manipulatives to enhance mathematics achievement. They found that using manipulatives was effective for improving mathematic outcomes. Manipulative type (i.e., concrete or virtual) did not have moderating effects. Similar to Methe et al., Burns et al., and Shin et al., this study did not address specific instructional practices. As such, it is not clear as to what other instructional practices could be contributing to the treatment effects.

Intervention Implementer and Duration. The role of intervention implementer in moderating intervention effects was only evaluated in Peltier et al. and Lei et al. Peltier found no significant differences in mathematics outcomes when the intervention was implemented by a researcher or a teacher. In contrast, Lei et al. found that interventions implemented by teachers showed a larger effect size than interventions implemented by a researcher, and both a researcher and a teacher. They found no difference for interventions implemented in a one-on-one or small group setting. In addition, intervention duration was coded into two categories: 10 hours or less, or more than 10 hours. Results indicated that interventions with a total duration of fewer than 10 hours led to greater treatment effects. Methe et al. also coded for duration of intervention. The authors multiplied the reported duration of each intervention session by the number of total sessions. However, due to inconsistencies in reporting total duration of intervention across studies, the effect of intervention duration was not clear.

Intervention Fidelity. Methe et al evaluated treatment integrity using three criteria: (a) independent variable reliability, (b) percent of intervention sessions in which integrity assessments were conducted, and (c) method by which integrity assessments were conducted. All studies reported treatment integrity exceeding 90%. However, this variable was not assessed as a potential moderator. Shin et al. (2020) used the CEC 2014 Quality Indicators to evaluate the intervention fidelity of the studies in their meta-analyses. They found that studies meeting all quality indicators of implementation fidelity (i.e., adherence, dosage, regularity) resulted in a greater effect size. In Peltier et al. (2020) analyses were not conducted to investigate the role of intervention fidelity due to insufficient information provided by the articles. Lastly, Burns et al. and Lei et al. did not examine intervention fidelity.

Methodological Quality

Evaluating the rigor of a study is necessary to provide evidence for internal validity. This is especially important when using SCD methodology, which does not employ a traditional control group. That being said, methods for measuring the quality of studies vary. All prior meta-analyses evaluated the quality of studies, except one (Burns et al., 2010). To begin with, Methe (2012) used visual analysis to determine the experimental control of the studies included in their meta-analysis. Experimental control was determined through visual analysis and was quantified on a continuous 10-point scale. In order to earn all 10 points, the study must show prediction, affirmation of the consequent, verification, and replication. The authors reported that 70% of interventions received experimental control ratings of 8 or above. However, there was a negative relationship between experimental control and treatment effects. According to the authors, most studies did not have evidence of verification or replication.

Peltier et al. (2020) created a coding sheet that aligned with the What Works Clearinghouse (WWC) Design Standards (Kratochwill et al., 2010). This was used to evaluate the studies. Studies were evaluated against the following criteria: (a) the independent variable was systematically manipulated, (b) interobserver agreement (IOA) data were reported, (c) IOA were collected on a minimum of 20% of baseline and intervention phases, (d) IOA scores met minimum quality, (e) there were a minimum of three attempts to demonstrate treatment effects at three different points in time, and (f) there were a minimum of three data points in baseline and intervention phases for multiple-baseline and multiple-probe experiments or a minimum of four data points for alternating treatment designs. It was reported that 11 studies met standards without reservations, 23 studies met standards with reservations, and 15 studies did not meet standards. Findings indicated that study quality did not moderate effect sizes. Lei et al. also assessed quality of study using the WWC Design Standards (Kratochwill et al., 2013). Each study was evaluated against five standards and resulted in one of three designations: (a) met standards without reservations, (b) met standards without reservations, or (3) did not meet standards. Unlike Peltier et al., the authors did not assess quality of study as a potential moderator, and instead excluded the studies that did not meet standards from their analyses (n = 2).

Shin et al. evaluated study quality using the CEC 2014 Quality indicators. A total of 22 indicators were assessed across eight categories: (1) context and setting, (2) participants, (3) intervention agents, (4) descriptions of practice, (5) implementation fidelity, (6) internal validity, (7) outcomes measures and dependent variables, and (8) data analyses. Each study received a score from 0 to 22. Altogether, the 20 studies had an average quality score of 20.56. Results indicated that study quality did not significantly moderate the treatment effect.

Effect Sizes

There are many established effect sizes in single-case design methodology. As a result of this, studies not only use different effect sizes, but also report multiple effect sizes. For example, Burns et al. (2010) used PAND, which was then converted to phi, a more robust effect size (Parker, Vannest, Brown, 2009). In contrast, Methe (2012) used improvement rate difference (IRD), and index that examines data nonoverlap between phases. IRD is compatible with visual analysis, can easily be calculated and interpreted and has stronger sensitivity than PND (Parker, Vannest, Brown, 2009). However, one limitation of IRD is that it has low discriminability of large effect sizes, such that when there is no overlap between phases, IRD assigns an effect size of 1.0. On the other hand, PND is an overlap-based index and thus has a ceiling effect deficiency. Shin et al. (2020) computed effect sizes using a two-level multilevel model. Level 1 examined within-case changes, and Level 2 between-case variances. A between-case standardized mean difference was calculated (BC-SMD). Peltier et al. (2020) used both Tau-*U*, a non-overlap index, and BC-SMD. Lei et al. (2020) only reported Tau-*U*.

Although there are many established effect size estimates that are commonly used in single-case design research (SCD), many of these effect size estimates pose significant limitations. For instance, when combining results from SCD studies and when comparing results from SCD research to that of group design research, effect size estimates must be standardized and take into account between-case variation (Shadish et al., 2015). As a result of this, researchers can combine results from studies that have distinct methodologies, such as studies with different outcome measures. However, many of the non-overlap effect sizes (e.g., PAND, PNF, IRD, Tau-*U)* which have been used in prior meta-analyses of SCD research do not take into account between-case variation. Additionally, many of these effect sizes (i.e., PAND, PND) do not take into account the trend and autocorrelation that exists in single-case data. As such, in the present meta-analyses I computed between-case standardized mean difference effect size estimates, which not only take into account within and between case variation, as well as autocorrelation and trend, but also include a confidence interval and standard error.

Research Questions

1. What is the effectiveness of mathematics interventions for students with mathematics learning disabilities (LD) and mathematics difficulties (MD) in grades K-5?
2. Does intervention effectiveness vary as a function of participant characteristics (i.e., grade level, MD/LD status)?
3. Does intervention effectiveness vary as a function of intervention characteristics (i.e., intervention implementer, intervention duration, intervention type, group size, mathematics domain).
4. Does intervention effectiveness vary as a function of study quality?

Chapter 2

Method

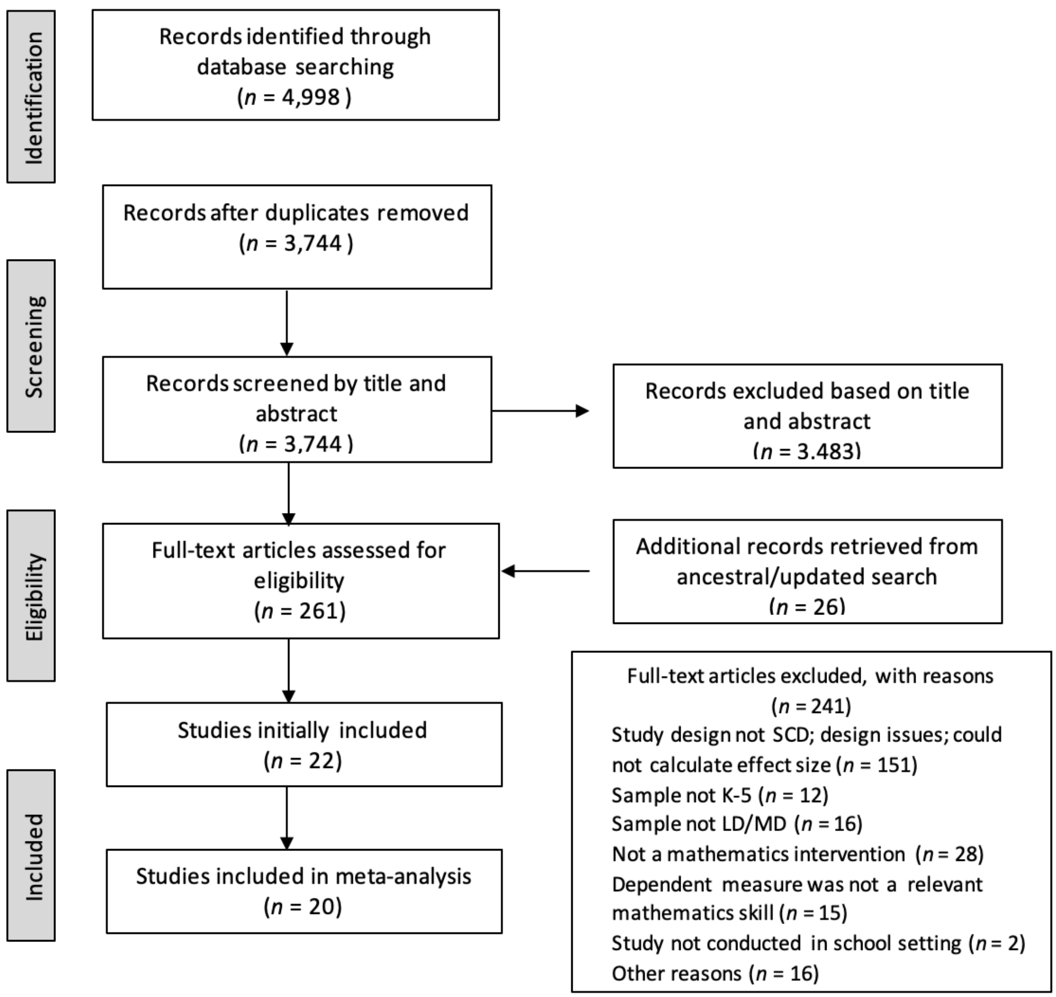
In this section, I first describe the search criteria and screening procedures for identifying the intervention studies that will be included in the meta-analysis. Second, I describe the procedures for coding the selected studies. Lastly, I will describe the data analysis plan for calculating effect sizes and testing for moderators.

Search Procedures and Criteria

The current study aims to synthesize data obtained from studies on mathematics interventions for students in grades K-5 with LD or MD. An initial search was conducted in July 2020 and updated on July 2021. To identify studies, I first conducted multiple searches in *Education Resource Information Center* (ERIC), *PsycInfo*, and *ProQuest* using a combination of the following search terms and roots: math* AND learning disabilit* OR learning difficult* OR learning problem* OR at risk OR mathematics difficult* OR mathematics disabilit* AND intervention OR instruction OR remediation OR tutoring OR training. The search was limited to academic articles and dissertations published from 2000-2021. The year 2000 was selected as the starting point to coincide with the reform movement of the National Council of Teachers of Mathematics, which sought to improve the quality of mathematics instructional programs for students. Moreover, the early 2000's marked the beginning of the No Child Left Behind Act which mandated and held schools accountable for providing effective instruction to all students. Studies identified using these search processes were imported into an Excel file where duplicates were eliminated. A total of 4,998 studies were identified through the online searches, and 3,744 remained after duplicates were removed. The screening process is described below and outlined in Figure 1.

Figure 1

Flow Chart of the Search and Screening Process



Title and Abstract Screening

After duplicates were removed, the remaining 3,744 studies were screened by title and abstract using the inclusion criteria below.

1. The study evaluated the effectiveness of a mathematics intervention.
2. The sample included students with a mathematics disability or learning disability in Grades K-5.
3. The study was conducted using a single-case design (SCD) that allowed for the calculation of between case standardized mean difference effect sizes (i.e., multiple baseline, multiple probe, and reversal designs).
4. The study was published in English.

Studies that remained following the title and abstract screening were downloaded for a full-text screening. If the title and abstract of a study did not report the information outlined in the inclusion criteria, then the study was retained for full-text review.

Ancestral Search. I also conducted an ancestral search by reviewing studies and references within prior published meta-analyses (i.e., Burns et al., 2010; Gersten et al., 2009; Shin et al., 2020; Stevens et al., 2018; Zhang & Xin, 2012). Furthermore, a hand search for studies by authors who conduct mathematics research was conducted in major journals in special education (i.e., *Journal of Learning Disabilities, Learning Disabilities Research & Practice, Learning Disabilities Quarterly, Exceptional Children, Remedial & Special Education, and Journal of Special Education).* The titles and abstracts of these studies were examined. A total of 26 studies were identified through the ancestral search and retained for full-text examination.

Full-Text Screening

Following the title/abstract screening and the ancestral/hand search, a total of 261 studies were downloaded for in-depth examination. In addition to the criteria used in the title and abstract screening procedure, the studies had to meet the following criteria:

1. The study included at least one outcome measure assessing mathematics performance.
2. Studies reported data needed to calculate effect size. That is, the study must include a graph depicting the results of the study. Additionally, the study must include at least three participants, and a minimum of three data points per baseline and intervention phase.
3. The intervention was implemented in a school setting as part of the school day program. (i.e., not at home, clinic, or camp)
4. The study was conducted in the United States.

Exclusionary Criteria. In addition to meeting the inclusion criteria, studies were excluded if they met specific exclusion criteria outlined below:

1. Participants had comorbid disabilities (e.g., Autism, Attention Deficit Hyperactivity Disorder, Emotional Disorder).
2. The goal of the intervention was to change student behavior (e.g., behavior contingency intervention) and did not target the remediation of a mathematics skill.
3. The goal of the intervention was to only improve fluency in mathematics, without providing participant with instruction on a specific mathematical skill.

4. The study was a published dissertation, a master's thesis, a case study, or was not an empirical study.

A total of 20 SCD studies met criteria for inclusion in the meta-analysis.

Coding of Studies

A coding sheet was developed to record the relevant features of the 20 articles that met criteria for inclusion (see Appendix A). The coding sheet was piloted and modified to ensure that the items accurately captured all relevant features of the studies. Each study was read and coded for the following: (a) participant characteristics, (b) study design characteristics, (c) intervention type and characteristics, and (d) methodological quality. The coding process was iterative and involved coding the articles, reviewing the codes with an expert in the field, revising the coding scheme, and re-coding the articles. Data was coded for the following study characteristics and used for descriptive purposes and/or examined as potential effect size moderators.

Participant Characteristics

Demographics. Demographic participant information for each study was coded and included: (a) grade level, (b) sex, (c) ethnicity/race, (d) English Language Learner (ELL) status, (e) eligibility for free or reduced lunch (FRL), (f) cognitive ability and (g) mathematics achievement. Grade level was initially coded as (1) Grade K-2, (2) Grade 3-4, (3) Grade 5, or (4) mixed. However, given the small number of studies with students at each grade level, these categories were collapsed for the statistical analyses. Final code categories were: (1) Grade 2, (2) Grade 3, or (3) Grade 4-5. Grade 2 was a group of its own because no studies included students in Kindergarten or Grade 1. Additionally, Grade 4 and Grade 5 were combined into one group to make the levels comparable in size for the statistical analyses. It is noted that prior research has coded grade level in similar ways, differentiating between students in the lower, mid, and upper grade levels (Lei et al., 2020; Lein et al., 2020). Separating students based on grade level is also often examined as a moderator because mathematics instruction generally looks different across the years. This is especially true following the Common Core State Standards released in 2010, which outline standards and provide guidelines of what students should be learning at every grade level from Kindergarten to Grade 12.

Regarding ethnicity, ELL status, eligibility for FRL, cognitive ability and achievement scores, I first coded whether or not the authors reported the specified information. For ethnicity/race, I recorded the percentage of students that fell within the following categories: White, African American/Black, Hispanic, Asian, or Other. For ELL status, I coded the percentage of students who were considered ELLs. For FRL eligibility, I recorded the percentage of students eligible for FRL. Furthermore, I recorded participants' scores for cognitive ability and achievement, if reported.

LD/MD Status. In addition, I recorded the type of criteria used to establish LD/MD status. First, I identified and coded the number of participants in each study that were identified as having LD and the number of students identified as having MD. Second, I identified and recorded the type of criteria used to establish LD/MD status: (1) specific cut score on a standardized test, researcher developed screening measure, or benchmark assessment, (2) teacher nomination, or (3) combined.

Intervention Characteristics

Each study was coded for specific intervention contextual characteristics including instructional setting, intervention implementer, group size, instructional duration, intervention fidelity, and intervention type. The instructional setting in each study was coded as (1) general education classroom, (2) special education/resource classroom, (3) other (e.g., library, computer lab, hallway), or (4) not reported. The intervention agent was coded as: (1) researcher, (2) teacher, (3) teacher and researcher, (4) other (e.g., paraprofessional), or (5) not reported. Group size was initially coded as: (1) individual/one-on-one, (2) pairs, or (3) group. However, given the small number of studies that provided instruction in pairs or a group, the two levels were combined as to make the groups more comparable for the statistical analyses. Final coding for group size included two levels: (1) individual/one-on-one or (2) small group.

The instructional time was calculated as total minutes of instruction and was coded as: (1) less than or equal to 600 minutes (i.e., 10 hours), or (2) greater than 600 minutes. The rationale for coding instructional time is based on research indicating that students with MD and LD who receive more than 10 hours of instruction are more likely to show improvement in mathematics, compared to students who receive less than 10 hours (Jitendra et al., 2018). If total instructional time was not reported, I calculated the average total instructional time by multiplying the total number of lessons/sessions by the average duration of each lesson/session. It is noted that intervention duration could not be coded as a continuous variable given that various studies did not clearly indicate the duration of the intervention. Studies reported a range for the number of lessons or for the minutes of instruction that students received. For example, Xin et al. (2020) reported that students received intervention "four times a week, with each session lasting about 20-30 min, for an average of 28 sessions (range from 22 to 36 sessions across the four participants)" (p. 110). As such, it was not possible to conclude exactly how long each intervention treatment lasted and because of this, intervention duration could not be coded as a continuous variable.

Intervention type was initially coded as: (1) explicit instruction, (2) schema-based instruction (SBI), (3) concrete-representational abstract (CRA), (4) cognitive/meta-cognitive instruction or (5) other. However, given the small number of studies, categories were collapsed as the following for statistical purposes: (1) explicit instruction, (2) schema-based instruction or cognitive strategy instruction, (3) concrete-representational abstract (CRA), and (4) other. These categories represent evidence-based practices and/or instructional approaches that have been consistently studied in prior mathematics research (Bouck & Park., 2018; Fuchs et al., 2021; Gersten et al., 2009; Jayanthi et al., 2008). Given the small number of studies included in the present meta-analysis, schema-based instruction and cognitive strategy instruction were collapsed into one category. Prior research has demonstrated that schema instruction and cognitive strategy instruction are both effective approaches when teaching students with MD and LD, and both approaches share various similarities (Jitendra et al., 2013; Jitendra & Star, 2011; Fuchs et al., 2021; Swanson et al., 2013; Swanson et al., 2014). Schema-based and cognitive instruction both involve the use of cognitive processes (e.g., paraphrasing, predicting answer, evaluating answer) that allow students to develop a deeper understanding of mathematical concepts, and help students organize information so that they can solve the problem (Fuchs et al., 2021; Montague & Dietz, 2009; Montague et al., 2011). Strategies such as analyzing the problem, monitoring strategy use, and evaluating the solution are embedded both within schema-based instruction (Jitendra et al., 2013; Jitendra et al., 2016; Jitendra et al., 2015) and cognitive strategy instruction (Montague & Dietz, 2009; Montague et al., 2011). In short, both approaches involve the use of cognitive and metacognitive processes to arrive at a solution to the problem. However, unlike cognitive strategy instruction, schema-based instruction specifically focuses on teaching students how to identify and recognize the underlying problem structure (e.g., change, compare) and to represent problems using visual-schematic diagrams. Students learn to plan and execute the appropriate mathematical computation and to evaluate their answers. Thus, schema-based instruction also incorporates the use of cognitive strategies, but the primary focus is on teaching students how to represent problems.

Furthermore, fidelity of implementation and social validity were both coded as dichotomous variables. I coded whether fidelity of implementation was reported and indicated the fidelity score. Lastly, I recorded whether social validity was reported.

Dependent Variable and Outcome Measure

Characteristics of the dependent variable and outcome measure for each study were also coded. The dependent variable refers to the mathematics domain targeted by the intervention. The dependent variable was coded as one of the following: (1) addition/subtraction, (3) multiplication/division, and (3) fractions. The outcome measure for the dependent variable was coded as (1) digits correct per minute, (2) number of problems correct, or (3) percent of problems correct. These categories represent common measures for assessing mathematics competency. Lastly, I coded whether interrater agreement data was collected for the dependent variable and report the average percentage of agreement.

Study Design Characteristics

Each study was coded for the type of single-case design it used. Given limitations to calculating effect sizes for certain single case designs (e.g., changing-criterion treatment), studies were only included and coded in this meta-analysis if they used a variation of the following: withdrawal/reversal or multiple baseline/multiple probe across participants. No studies used a withdrawal/reversal design, and as such this design will not be discussed further.

Study Quality. The methodological quality of studies was evaluated using a coding sheet that was developed based on the What Works Clearinghouse (2020) guidelines (see Appendix B). WWC provides a guide to examine the methodological rigor of studies and examine evidence of experimental control or internal validity. According to WWC, studies can meet criteria for one of three ratings: meets standards without reservation, meets standards with reservation, or does not meet standards. Each study was evaluated against the following criteria: (1) the independent variable was systematically manipulated; (2) the dependent variable was measured systematically over time by more than one assessor; (3) inter-assessor agreement was collected in each phase and on at least 20% of the sessions in each phase (e.g., baseline, intervention), (4) inter-assessor agreement met threshold (.80 to .90 if measured by percentage agreement; 0.60

if measured by Cohen's kappa), (5) study includes at least 3 attempts to demonstrate an intervention effect at three different points in time or with three different phase repetitions (i.e., tiers within MBDs), and (6) depending on the design type, phases must meet criteria involving the number of data points. Failure to meet any of these criteria would result in a study rating of does not meet standards.

Multiple baseline and multiple probe. To meet standards without reservations, a multiple baseline or multiple probe study design must have a minimum of six phases with at least five data points per phase. To meet standards with reservations, there must be a minimum of six phases with at least three data points per phase. If any phase has fewer than three data points, this will result in a rating of does not meet standards.

Visual analysis. Following WWC recommendations from initial pilot standards, studies that met standards without or with reservations were further examined for evidence for demonstrating a functional (causal) relation (Horner et al., 2005; Kazdin, 2011). Visual analysis of (a) level, (b) trend, (c) variability, (d) overlap, (e) immediacy of effect, and (f) consistency of data patterns across similar phases was used to determine whether there was evidence of a functional relation. Level refers to the overall average (mean) performance during a phase (e.g., baseline, treatment). Trend refers to the tendency of performance to systematically increase or decrease over time (i.e., slope of the best-fit straight line for the dependent variable within a phase). Variability refers to the degree to which the data points fluctuate around a mean or slope in a phase. Overlap refers to the proportion of data from one phase that overlaps with data from the previous phase. Immediacy of effect refers to the change in the mean from one phase to the other (e.g., change in mean from baseline to treatment). Consistency of data refers to the extent to which there is consistency in the data patterns from phases with the same conditions. In addition to meeting these criteria, the study must have at least three demonstrations of the intervention effect. These criteria were assessed individually and collectively to determine if a functional relationship existed between the independent and dependent variables. Studies then met criteria for one of three ratings: strong evidence, moderate evidence, or no evidence.

Inter-rater agreement (IRA)

To ensure reliability of the coding process, studies were coded by the first author and two additional raters. The first author served as the first rater and coded all studies. Two trained undergraduate students served as the second raters. Before beginning the IRA and coding process, the first author met with the second raters for a 30-minute training session before each coding phase. During the training sessions, the first author and raters coded one study together. The first author reviewed the coding criteria and answered clarifying questions. Following the trainings, the first author and coders independently practiced coding three excluded studies until it was determined that interrater agreement was at least 90%. Inter-rater agreement occurred at four different time points: (1) title and abstract screening, (2) full-text screening, (3) coding of studies, and (4) data extraction. The first author completed all coding at each step. The second raters independently coded 25% of randomly selected studies at each step. One rater completed coding for the title and abstract screening and the data extraction, while the second rater completed the coding for the full-text screening and coding of studies. In instances of disagreement, raters discussed the specific items of disagreement until they reached a consensus on the item. Changes to the coding criteria were also made as necessary.

IRA was calculated with kappa coefficients (Cohen, 1965) using the following formula: $(P_o – P_c)/(1- P_c)$, where P_o is the proportion of agreements between observers on occurrences and non-occurrences and P_c is the proportion of expected agreements on the basis of chance. All inter-rater agreement scores met minimum quality thresholds of .60 or higher. IRA resulted in the following kappa coefficients: .81, .95, and .96 for title and abstract screening, full-text screening, and data extraction, respectively. Each article was coded for participant characteristics, intervention characteristics, WWC standards, and evidence for a functional relation, and IRA resulted in the following kappa coefficients: 1.00, .82, and .89, and .89, respectively. In instances of disagreement, raters discussed the specific items of disagreement until they reached a consensus on the item.

Data Analyses

Descriptive statistics were first conducted to understand the characteristics of the sample. The distribution of effect sizes were examined for evidence of any unusual patterns in the data (e.g., outliers) using a forest plot of the effect sizes, as well by graphing the effect sizes on a boxplot. A random-effects model was used because of the realistic assumption that effect sizes are sampled from a population of effects that likely vary due to variation in study treatment conditions, quality of measures, etc. (Borenstein et al., 2009).

Effect Size Calculation and Analytic Method

Standardized effect sizes are commonly used to synthesize and summarize the treatment effects of an intervention. Standardized effect sizes allow researchers to compare intervention results across studies with varying outcome measures, which is important when synthesizing the evidence base to establish evidence-based practices. Although it is common in group design methodology (e.g., randomized experiment) to use standardized effect sizes to evaluate the effectiveness of an intervention, SCD research has used varying methods to evaluate intervention effects, including parametric and nonparametric methods. For instance, overlap statistics (e.g., percentage of nonoverlapping data, percentage of data exceeding the mean) improvement rate difference (IRD), and Tau-U, in addition to visual analysis, are often used to summarize results from SCD studies (Shadish et al., 2014). However, new methods for computing effect sizes in SCD research continue to be refined and developed.

Hedges, Pustejovsky, and Shadish (2012, 2013) and Pustejovsky, Hedges, and Shadish (2014) proposed an effect size for single-case studies that is comparable to the standardized mean difference from a between-subjects randomized experiment. The between-case standardized mean difference (BC-SMD) effect size is analogous to Cohen's *d* and, when corrected for small sample bias, is equivalent to Hedges' *g* (Shadish et al., 2014). Unlike other effect sizes, the BC-SMD statistic takes into account both within-case and between-case variation; this is important to compare effect sizes across studies. Effect sizes that are only standardized within-case cannot be compared to studies with effect sizes estimated using between-case variation, as they estimate different parameters. (Shadish et al., 2008; Shadish et al., 2014)

The BC-SMD effect size estimation approach models data using a hierarchical model that takes into account the nested structure of SCD data (Valentine et al., 2016). Effect sizes are calculated using a two-level model, where Level 1 is the within-case variation for every individual case (i.e., variation within participants), and Level 2 is the between-case variation (i.e., variation across participants). The BC-SMD effect size index takes into account across-participant variation. As a result, the effect size can only be applied to studies that include multiple individuals. That is, the study must include at least three individuals in order to examine individual differences across participants. As such, the BC-SMD effect size can only be estimated for studies with the following single-case designs when at least three participant-level cases are utilized: treatment reversal/withdrawal, multiple baseline across-participant, and multiple-probe across participant.

In the present meta-analyses, BC-SMD effect sizes were estimated to measure intervention effects. BC-SMD effect sizes were estimated using the single-case design hierarchical linear model (scdhlm) package in R (Pustejovsky et al., 2021). Graphed data for each outcome measure in the studies was used to compute effect sizes. Images of each graph were uploaded to the WebPlotDigitizer (Rohatgi, 2020), a web-based application that extracts baseline and intervention data. This program assisted the author in plotting the graph coordinates and data points and creating excel files with the baseline and intervention data. Data extraction was conducted independently by the first author and a second rater. The second rater extracted data for 25% of the studies to assess for reliability in the coding. IRA resulted in a kappa coefficient of 0.96 In instances of disagreement, raters discussed the specific items of disagreement until a consensus was reached.

Meta-analytic models were estimated using the robumeta R package (Fisher & Tipton, 2015; Fisher et al., 2017). Studies with multiple outcome measures and those reporting more than one effect size (correlated effects) were included in the meta-analysis. As such, this meta-analysis used robust variance estimation (RVE; Hedges et al., 2010; Tanner & Tipton, 2014; Tanner et al., 2016) to address issues of dependency in effect sizes. RVE requires the mean correlation between all pairs of effect sizes within a study (*rho*) to estimate study weights and between-study variance. Because correlations between the effect sizes in each study is usually not reported, a sensitivity analysis was conducted in which varying *rho* values (0.2, 0.5, and 0.8) were used in the meta-analytic models to assess whether the between-study variance is robust across different estimates of the within-study correlation between effect sizes (ρ).

In all meta-regression analyses, random effects models weighted by the inverse variance of effect sizes were used. Studies that yield a more precise estimate of the effect size were assigned more weight. Tanner-Smith and Tipton (2014) recommend choosing a weighting method based on the most common source of dependency in the data structure. A correlated effects model will be appropriate if a majority of the dependency in the data arises from measurements made on the same number of subjects. In other words, multiple effect sizes were estimated based on the same participant samples. In contrast, a hierarchical effects model will be more appropriate if most of the data dependency arises from multiple studies being nested within a larger cluster, such as a laboratory or research group. In the present meta-analysis, RVE with correlated effects was used to model the data.

Evaluating Heterogeneity of Effect Sizes and Moderator Analyses. Heterogeneity of effect sizes was evaluated using the I^2 statistic, which indicates the extent to which variation in effect sizes is likely due to potentially explainable variation rather than sampling error (chance). Large I^2 values suggest that most variation in effect sizes is between-studies, and thus conducting moderator analyses is critical. The following guidelines were used for interpreting the I^2 statistic: I^2 = 25% suggests a small amount of heterogeneity, I^2 = 50% suggests medium heterogeneity and I^2 = 75% suggests large heterogeneity (Copper et al., 2009; Higgins et al., 2003).

A multiple meta-regression analysis was conducted to explore the effects of moderators (i.e., MD/LD status, grade level, interventionist, instructional group size, duration of intervention, intervention model type, mathematics domain, and study quality) on students' mathematics outcomes. Moderator analyses only included studies that provided sufficient data to code the moderator variables. Meta-regression allows the inclusion of multiple moderators in the model.

Publication bias. Publication bias may be a threat to the validity of findings because of the possibility that unpublished studies with nonsignificant findings have been excluded. To evaluate the likelihood of publication bias as a threat within the present meta-analysis, I first visually inspected the symmetry of a funnel plot illustrating the relation between effect size and study precision (standard error of the effect estimate) and used the Egger's statistical test (Egger et al., 1997) to assess funnel plot asymmetry. I supplemented this analysis with the Duval and Tweedie's (2000) trim-and-fill procedure that is used to estimate the number of studies missing from a meta-analysis and produces a funnel plot that is more symmetric.

Chapter 4

RESULTS

In this section, I first report descriptive statistics for the sample of participants, study characteristics, outcome measures, and intervention characteristics. Next, I report the overall effect of mathematics interventions on students with specific learning disabilities in mathematics (LD) and mathematics difficulties (MD). Lastly, I report results from the meta-regression and describe moderator effects.

Descriptive Statistics

A total of 22 studies and 34 effect sizes initially met eligibility criteria for the meta-analysis. All effect sizes were plotted on a forest plot, boxplot, and histogram to review their distribution. After examining the distribution of effect sizes, 2 studies and 3 effect size estimates with the following magnitudes were removed as outliers: 24.72, 6.41, and 5.01 (Bouck et al. 2021, Mancl et al. 2012, Morin et al. 2017). When closely examining the 3 effect sizes that were removed as outliers, it was determined that the estimates were significantly large due to floor and ceiling effects. Scores for participants in all 3 studies were consistently at zero during the baseline phase. During the treatment phase, participants demonstrated immediate improvements, with many of the participants showing ceiling effects.

It is also noted that 3 effect sizes were removed prior to analyses given that they did not meet WWC standards (Kim et al., 2015) or had weak evidence for a functional relation (McKevett et al., 2019, Preston, 2016). According to single-case study design guidelines, effect sizes should only be calculated for studies that meet design criteria and demonstrate moderate or strong evidence of a functional relation (Kratochwill et al., 2013; Kratochwill et al., 2021; WWC, 2014). As such, the results that follow were conducted without the effect sizes that were outliers and the effect sizes that did not meet WWC standards and demonstrated weak evidence of a functional relation. Results thus included 20 studies with a total of 31 effect size estimates (see Figure 2).

Publication Bias

Publication bias was first assessed by visually inspecting the asymmetry in a funnel plot of the effects size estimates against their standard errors. An Egger's regression test was then conducted to assess for significant publication bias. Lastly, Duval and Tweedie's (2000) trim-and-fill of the funnel plot was conducted to correct any funnel plot asymmetry. Visual inspection of the funnel plot suggested publication bias against studies with smaller sample sizes, as well as studies with smaller effect sizes. Given that statistically significant findings are more likely to be published compared to null or negative findings (Rosenthal, 1979; Shadish et al., 2016), small and nonsignificant findings tend to be disproportionally underrepresented in meta-analyses (Gage et al., 2017). Consistent with visual analysis of the funnel plot, the Egger's test showed significant publication bias ($z = 3.58$, $p < .001$). Findings from this meta-analysis thus indicate the existence of publication bias and warrant caution when interpreting findings.

Figure 2

Forest Plot of Effect Sizes with 95% CI

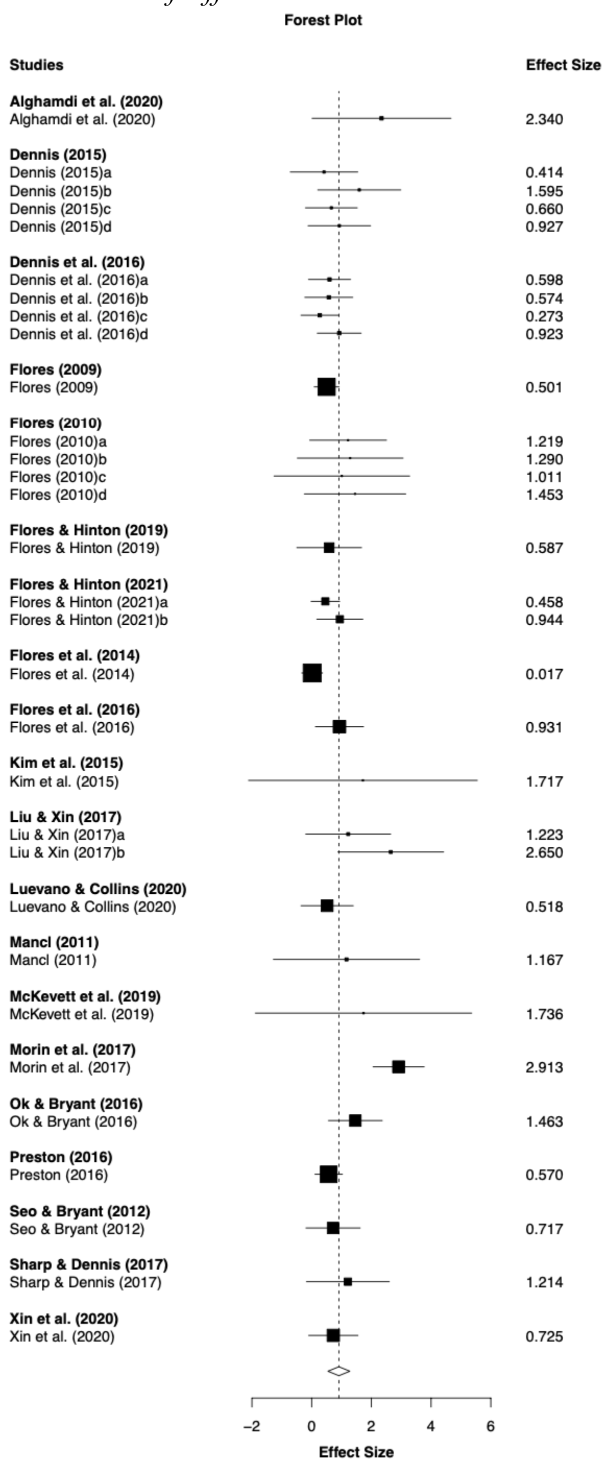
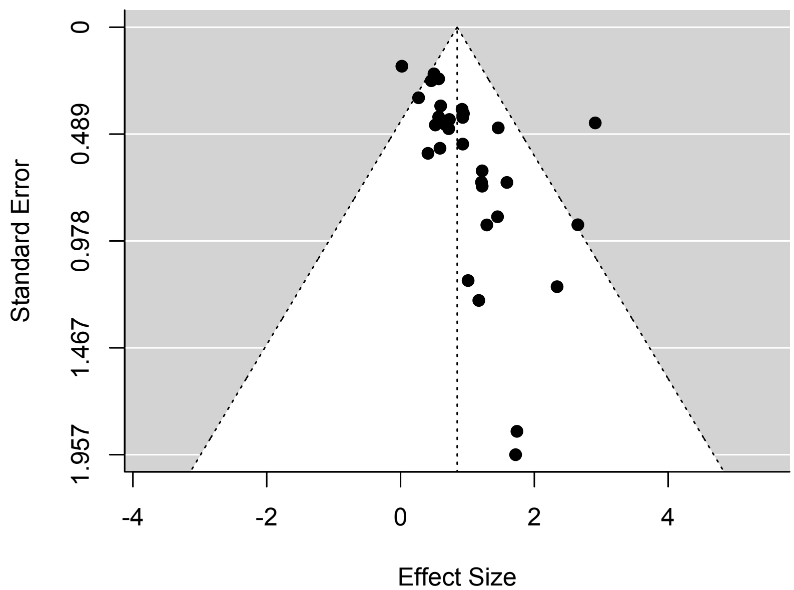


Figure 3

Funnel Plot of Standard Errors by Effect Size



Participants

As can be seen in Table 1, a total of 88 participants were represented across the studies, with 44% of participants being female. In terms of grade level, 30% of participants were in second grade, 34% third grade, 16% fourth grade, and 20% fifth grade. No studies included students in kindergarten or first grade. In terms of disability status, 39% of participants had a specific learning disability in mathematics (LD) and 61% were demonstrating mathematics difficulties (MD). Regarding race/ethnicity, 25% were African-American/Black, 44% Hispanic/Latino, 24% White, 6% Asian, and 1% Mixed. Cognitive ability was only reported for 40% of participants, or 9 studies. Standard scores for cognitive ability ranged from 71 to 133, with a mean of 97.6 and a median of

96. Likewise, mathematics academic achievement was reported for 66% of participants or 13 studies. However, score reporting varied with some studies reporting standard scores, while others reported percentiles or scaled scores. Regarding English learner status, 23% of participants were identified as non-English learners and 16% as English learners. Data was not reported for 61% of participants, or 12 studies. Lastly, regarding eligibility for free or reduced lunch, 19.5% of participants received free or reduced lunch and 3.5% did not receive free or reduced lunch. Free or reduced lunch data was not reported for 77% of participants.

Intervention Characteristics

As can be seen in Table 1, intervention type varied across studies, with 39% of studies examining the effectiveness of the concrete-representational-abstract (CRA) model, 29% explicit/direct instruction, 22% schema/cognitive instruction, and 10% other (i.e., extensive practice, independent practice, game-based learning). For statistical purposes, it is noted that studies were coded such that the intervention type represented the primary form of intervention implemented in the study. For example, Dennis (2015) primarily provided students with explicit instruction and only used the CRA approach as needed. In this case, the intervention model was coded as explicit instruction. Likewise, all studies that used various forms of the CRA model were collapsed into one category. For example, Flores and Hinton (2021) examined the effectiveness of the concrete-representational abstract integrated model (CRA-I), whereas Flores et al. (2014), Flores et al. (2016), and Flores and Hinton (2019) examined the effectiveness of the combination of the concrete-representational-abstract sequence and the strategic instruction model

(CRA-SIM). In all these instances, the intervention model was coded as CRA. Furthermore, schema-based instruction and cognitive/metacognitive instruction were collapsed into one category, given the overlap between both intervention models.

Interventions were implemented primarily by researchers (68%), followed by teachers (23%) and computer (9%). Interventions were implemented in the general education setting (10%), resource or special education classroom (19%) and in other settings (71%). Other settings included library, computer lab, and conference room. Intervention duration was either less than 600 minutes (65%) or more than 600 minutes (26%). Nine percent of studies did not report intervention duration. Interventions were delivered individually (74%) or in a small group (26%). A small group included 2 or more students.

Table 1

Summary of Participant and Intervention Characteristics

	Participants			Intervention				
Study	Number; Gender	Grade	Disability	Type	Duration (min)	Interventionist	Math Domain	Treatment Fidelity
Alghamdi et al. (2020)	*n* = 3; 0 F	5	100% LD	Schemas/Cog	less than 600	Researcher	Multiplication/Division	94%
Dennis (2015)a	*n* = 3; 1 F	2	100% MD	Explicit	more than 600	Researcher	Addition/Subtraction	NR
Dennis (2015)b	*n* = 3; 1 F	2	100% MD	Explicit	more than 600	Researcher	Addition/Subtraction	NR
Dennis (2015)c	*n* = 3; 3 F	2	100% MD	Explicit	more than 600	Researcher	Addition/Subtraction	NR
Dennis (2015)d	*n* = 3; 1 F	2	100% MD	Explicit	less than 600	Researcher	Addition/Subtraction	97%
Dennis et al. (2016)a	*n* = 3; 1 F	2	100% LD	Explicit	more than 600	Researcher	Addition/Subtraction	92%
Dennis et al. (2016)b	*n* = 3; 1 F	2	100% LD	Other	more than 600	Researcher	Addition/Subtraction	92%
Dennis et al. (2016)c	*n* = 3; 3 F	2	100% LD	Other	more than 600	Researcher	Addition/Subtraction	92%
Dennis et al. (2016)d	*n* = 3; 3 F	2	100% LD	Explicit	more than 600	Researcher	Addition/Subtraction	92%
Flores (2009)	*n* = 6; 2 F	3	33% MD	CRA	less than 600	Researcher	Addition/Subtraction	100%
Flores (2010)a	*n* = 3; 0 F	3	100% MD	CRA	less than 600	Teacher	Addition/Subtraction	100%
Flores (2010)b	*n* = 3; 0 F	3	100% MD	CRA	less than 600	Teacher	Addition/Subtraction	100%
Flores (2010)c	*n* = 3; 1 F	3	100% MD	CRA	less than 600	Teacher	Addition/Subtraction	100%
Flores (2010)d	*n* = 3; 1 F	3	100% MD	CRA	less than 600	Teacher	Addition/Subtraction	100%
Flores & Hinton (2019)	*n* = 3; 1 F	3	100% MD	CRA	less than 600	Researcher	Multiplication/Division	100%
Flores & Hinton (2021)a	*n* = 5; 1 F	2	100% MD	CRA	less than 600	Teacher	Addition/Subtraction	95%

Table 1 (continued)

Study	Number; Gender	Grade	Disability	Type	Duration (min)	Interventionist	Math Domain	Treatment Fidelity
Flores & Hinton (2021)b	*n* = 5; 1 F	2	100% MD	CRA	less than 600	Teacher	Addition/Subtraction	95%
Flores et al. (2014)	*n* = 4; 1 F	4-5	100% LD	CRA	less than 600	Researcher	Multiplication/Division	100%
Flores et al. (2016)	*n* = 3; 1 F	3	100% MD	CRA	less than 600	Researcher	Addition/Subtraction	95%
Kim et al. (2015)	*n* = 3; 1 F	4	100% MD	CRA	less than 600	Researcher	Fractions	95%
Liu & Xin (2017)a	*n* = 3; 1 F	4	100% LD	Explicit	less than 600	Researcher	Multiplication/Division	96%
Liu & Xin (2017)b	*n* = 3; 1 F	4	100% LD	Explicit	less than 600	Researcher	Multiplication/Division	96%
Luevano & Collins (2020)	*n* = 4; 2 F	2	100% MD	Schemas/Cog	less than 600	Researcher	Addition/Subtraction	96%
Mancl (2011)	*n* = 3; 1 F	4-5	100% LD	CRA	less than 600	Teacher	Addition/Subtraction	100%
McKevett et al. (2019)	*n* = 3; 3 F	5	100% MD	Other	less than 600	Computer	Fractions	99%
Morin et al. (2017)	*n* = 6; 5 F	3	83% MD	Schemas/Cog	less than 600	Researcher	Addition/Subtraction	NR
Ok & Bryant (2016)	*n* = 4; 2 F	5	100% LD	Other	less than 600	Researcher	Multiplication/Division	98%
Preston (2016)	*n* = 6; 3 F	4-5	83% MD	Schemas/Cog	less than 600	Researcher	Multiplication/Division	98%
Seo & Bryant (2012)	*n* = 4; 1 F	2-3	100% MD	Schemas/Cog	less than 600	Computer	Addition/Subtraction	96%
Sharp & Dennis (2017)	*n* = 3; 2 F	4	100% LD	Schemas/Cog	less than 600	Researcher	Fractions	98%
Xin et al. (2020)	*n* = 4; 2 F	3	100% MD	Schemas/Cog	more than 600	Computer	Addition/Subtraction	NR

Note. MD = mathematics difficulties; LD = specific learning disability in mathematics; NR = not reported

Study Characteristics

A total of 31 BC-SMD effect sizes were extracted from 20 studies. Studies were published during the years 2009 to 2021. Two of the studies were unpublished dissertations and 18 were published in peer-reviewed journals. As can be seen in Table 2, 84% of studies met WWC standards with reservations and 16% met standards without reservations. Of those studies that met WWC standards with and without reservations, 55% demonstrated moderate evidence of a functional relation and 45% demonstrated strong evidence. Sixteen percent of effect sizes used a multiple-baseline design and 84% used a multiple-probe design. All studies assessed effects across participants. The average number of baseline sessions was 4 across studies. The average number of treatment session was 10.5 across all studies. Although 97% of studies reported that they measured treatment integrity, only 84% of studies reported treatment integrity scores, which averaged 97% across studies. All studies measured interobserver agreement (IOA). The average IOA was 98% across all studies. In addition, 74% of studies measured maintenance outcomes. On average, 86% of students showed maintenance effects in each study. Maintenance follow-ups ranged from 1 to 6 weeks across the studies reporting maintenance outcomes. Social validity was measured in approximately 84% of studies.

Table 2
Study Characteristics and Quality

Study	Design	Type	WWC	Evidence	Shortest # of Baseline Sessions	Average # of Treatment Sessions	IOA (%)	Maintenance Measured
Alghamdi et al. (2020)	MP	Across participants	Meets w/Reservations	Strong	5	5	93	Y
Dennis (2015)a	MP	Across participants	Meets w/Reservations	Moderate	4	10	99	Y
Dennis (2015)b	MP	Across participants	Meets w/Reservations	Strong	4	10	99	Y
Dennis (2015)c	MP	Across participants	Meets w/Reservations	Strong	4	10	99	Y
Dennis (2015)d	MB	Across participants	Meets w/Reservations	Moderate	4	12	100	Y
Dennis et al. (2016)a	MP	Across participants	Meets w/Reservations	Moderate	5	16	97	N
Dennis et al. (2016)b	MP	Across participants	Meets w/Reservations	Moderate	5	16	97	N
Dennis et al. (2016)c	MP	Across participants	Meets w/Reservations	Strong	4	16	97	N
Dennis et al. (2016)d	MP	Across participants	Meets w/Reservations	Moderate	5	16	97	N
Flores (2009)	MP	Across participants	Meets w/Reservations	Strong	3	11	98	Y
Flores (2010)a	MP	Across participants	Meets w/Reservations	Strong	3	7	97	N
Flores (2010)b	MP	Across participants	Meets w/Reservations	Strong	3	5	97	Y
Flores (2010)c	MP	Across participants	Meets w/Reservations	Strong	3	6	97	Y
Flores (2010)d	MP	Across participants	Meets w/Reservations	Strong	3	10	97	N
Flores & Hinton (2019)	MP	Across participants	Meets w/Reservations	Strong	4	11	100	Y

Table 2 (continued)

Study	Design	Type	WWC	Evidence	Shortest # of Baseline Sessions	Average # of Treatment Sessions	IOA (%)	Maintenance Measured
Flores & Hinton (2021)a	MP	Across participants	Meets Standards	Moderate	5	8	100	Y
Flores & Hinton (2021)b	MP	Across participants	Meets w/Reservations	Moderate	5	8	100	Y
Flores et al. (2014)	MP	Across participants	Meets w/Reservations	Moderate	5	10	99	Y
Flores et al. (2016)	MP	Across participants	Meets w/Reservations	Moderate	5	13	100	Y
Kim et al. (2015)	MP	Across participants	Meets w/Reservations	Strong	3	10	100	Y
Liu & Xin (2017)a	MB	Across participants	Meets Standards	Moderate	5	10	93	Y
Liu & Xin (2017)b	MB	Across participants	Meets w/Reservations	Moderate	3	10	95	Y
Luevano & Collins (2020)	MP	Across participants	Meets w/Reservations	Moderate	3	10	96	N
Mancl (2011)	MP	Across participants	Meets w/Reservations	Moderate	3	17	100	Y
McKevett et al. (2019)	MB	Across participants	Meets Standards	Moderate	5	13	98	N
Morin et al. (2017)	MB	Across participants	Meets Standards	Moderate	5	12	91	Y
Ok & Bryant (2016)	MP	Across participants	Meets w/Reservations	Strong	3	15	99	Y
Preston (2016)	MP	Across participants	Meets Standards	Moderate	10	6	99	Y
Seo & Bryant (2012)	MP	Across participants	Meets w/Reservations	Strong	2	12	99	Y
Sharp & Dennis (2017)	MP	Across participants	Meets w/Reservations	Strong	6	6	100	Y
Xin et al. (2020)	MP	Across participants	Meets w/Reservations	Moderate	3	7	98	Y

Note. MP = multiple probe; MB = multiple baseline; Y = Yes; N = No

Outcome Measures

In terms of the dependent variables, 48% of individual studies targeted Number and Operations in Base Ten, 42% Operations and Algebraic Thinking, and 10% Number and Operations-Fractions. For the outcome measures, 26% of effect sizes used number of correct digits, 32% reported percent correct, and 42% calculated total correct raw scores. The outcome measures for 100% of the studies were researcher-developed. Regarding mathematics domains, 68% of studies examined interventions that targeted addition or subtraction, 22% examined multiplication or division interventions, and 10% examined fraction interventions.

Overall Effect of Mathematics Interventions

The overall weighted effect size estimate of mathematics interventions for students with MD and LD was 0.91 (95% CI = 0.54, 1.28, $p < .01$), indicating a significant positive effect of mathematics interventions. As can be seen in Table 3, the BC-SMD effect size estimates ranged from 0.02 to 2.91. All effect sizes showed positive effects of mathematics interventions. A sensitivity analysis was conducted to ensure that the overall effect size estimate was not sensitive across different estimates of the within-study correlations (ρ), which is used when calculating weights for the meta-analysis model. Sensitivity analysis indicated that the overall effect size estimate was relatively robust across different values of ρ.

Table 3

BC-SMD Effect Sizes

Study	BC-SMD	*SE*	CI	*df*	AC	ICC
Alghamdi et al. (2020)	2.34	1.19	[-0.85, 5.53]	4.39	0.64	0.37
Dennis (2015)a	0.41	0.58	[-1.12, 1.95]	4.48	0.88	0.00
Dennis (2015)b	1.59	0.71	[-0.12, 3.31]	6.28	0.77	0.00
Dennis (2015)c	0.66	0.44	[-0.54, 1.86]	4.26	0.22	0.74
Dennis (2015)d	0.93	0.54	[-0.58, 2.44]	3.83	-0.27	0.86
Dennis et al. (2016)a	0.60	0.36	[-0.22, 1.42]	8.59	0.26	0.25
Dennis et al. (2016)b	0.57	0.41	[-0.38, 1.52]	7.88	0.76	0.00
Dennis et al. (2016)c	0.27	0.32	[-0.52, 1.06]	6.04	0.87	0.00
Dennis et al. (2016)d	0.92	0.38	[0.08, 1.77]	9.35	0.10	0.30
Flores (2009)	0.50	0.21	[-0.01, 1.01]	6.51	0.36	0.89
Flores (2010)a	1.22	0.66	[-0.5, 2.94]	4.72	0.39	0.68
Flores (2010)b	1.29	0.91	[-1.4, 3.98]	3.43	0.23	0.78
Flores (2010)c	1.01	1.16	[-3.01, 5.03]	2.62	0.90	0.45
Flores (2010)d	1.45	0.87	[-1.03, 3.94]	3.72	0.94	0.00
Flores & Hinton (2019)	0.59	0.55	[-0.61, 1.78]	13.13	0.52	0.00
Flores & Hinton (2021)a	0.46	0.24	[-0.14, 1.05]	6.22	0.51	0.81
Flores & Hinton (2021)b	0.94	0.40	[0.11, 1.78]	17.43	0.60	0.16
Flores et al. (2014)	0.02	0.18	[-0.43, 0.47]	5.29	0.94	0.00
Flores et al. (2016)	0.93	0.41	[-0.07, 1.93]	6.32	-0.01	0.64
Kim et al. (2015)	1.72	1.96	[-5.2, 8.63]	2.54	0.30	0.83
Liu & Xin (2017)a	1.22	0.73	[-0.8, 3.25]	3.97	0.60	0.64
Liu & Xin (2017)b	2.65	0.90	[0.56, 4.74]	7.99	0.55	0.00
Luevano & Collins (2020)	0.52	0.45	[-0.42, 1.45]	19.84	0.52	0.04
Mancl (2011)	1.17	1.25	[-3.00 , 5.33]	2.78	-0.14	0.81
McKevett et al. (2019)	1.74	1.85	[-4.67, 8.15]	2.61	0.22	0.76
Morin et al. (2017)	2.91	0.44	[2.03, 3.8]	39.43	-0.08	0.17
Ok & Bryant (2016)	1.46	0.46	[0.47, 2.46]	13.35	0.68	0.00
Preston (2016)	0.57	0.24	[0.02, 1.12]	7.74	0.20	0.83
Seo & Bryant (2012)	0.72	0.46	[-0.42, 1.86]	5.94	0.76	0.00
Sharp & Dennis (2017)	1.21	0.71	[-0.73, 3.16]	4.12	0.66	0.66
Xin et al. (2020)	0.73	0.42	[-0.26, 1.71]	7.54	0.73	0.00

Note. SE = standard error; CI = confidence interval; AC = auto-correlation; ICC = intra-class correlation

Moderator Analyses

Heterogeneity of variance was evaluated using the I^2 statistic, which indicates the extent to which variation in effect sizes is likely due to sampling error (chance) versus potentially explainable variation. The I^2 statistic was 63.75%, indicating that a moderate amount of variation can be explained by differences across studies, as opposed to chance (Copper et al., 2009; Higgins et al., 2003). As such, a multiple meta-regression analysis was conducted to explore the effects of moderators and better understand the unexplained heterogeneity across the studies. Table 4 shows the results of the moderator analyses. For descriptive purposes, Table 5 shows a summary of the weighted averages of effect sizes for each moderator level. Weighted averages take into account the weight of each effect size, which was used when computing the overall effect size of mathematics interventions on students with LD and MD, as well as when conducting the moderator analyses. However, weighted averages do not take into account the effect of other variables (i.e., moderator variables), and as such they only provide supplemental descriptive findings from the meta-analysis and are interpreted along findings from the moderator analyses.

Participant Characteristics

Analyses were first conducted to determine whether participant grade level and disability status functioned as moderators. Grade level was coded into three categories: Grade K-2, Grade 3, or Grade 4-5, with Grade K-2 serving as the reference point. It is noted that no studies included students in kindergarten and first grade. As such, the K-2 category only included students who were in Grade 2. As a result of this, the moderator level will be referred to as Grade 2 from here on. Results indicated that grade level moderated intervention effects, such that effects for students in Grade 3 (β = 1.25, *SE* = 0.47, $p < 0.1$) were significantly higher than those in Grade 2. There was no significant difference in outcomes for student in Grade 2 and Grades 4-5 (β = 0.45, *SE* = 0.63, p = 0.51). However, it is noted that the degrees of freedom when comparing students in Grade 2 with students in Grade 4-5 were less than four and as such results may be considered unreliable (Fisher & Tipton, 2015). When conducting RVE analyses with the robumeta package in R software, the degrees of freedom are adjusted for small samples. Fisher and Tipton (2015) recommend that these corrections be implemented in all RVE analyses, and they are especially important when the number of studies is less than 40 and when covariates are unbalanced or highly skewed. As previously noted, weighted effect size averages were computed to aid in interpretation of results, along with the moderator analyses. For grade level, weighted averages ($\bar{X}_w$) were 0.67, 1.11, and 0.92 for students in Grade 2, Grade 3, and Grade 4-5, respectively. In line with results from the moderator analyses, students in Grade 3 showed, on average, higher effect sizes followed by students in Grade 4-5, and students in Grade 2.

Furthermore, disability status was assessed as a potential moderator. Disability was coded into two categories: specific learning disability in mathematics (LD) and mathematics difficulties (MD). Results indicated that disability status did not moderate intervention effects. Weighted averages ($\bar{X}_w$) were 0.97 and 0.82 for students with MD and LD, respectively.

Table 4
Moderator Analysis Results

Variables	Estimate (β)	*SE*	*t*	*df*	95% CI	*p* value
Grade Level						
Grade 3	1.25*	0.47	2.66	4.45	[-0.00-2.49]	0.05
Grade 4 to 5	0.45	0.63	0.72	3.67	[-1.35-2.25]	0.51
Disability						
MD	-0.03	0.51	0.07	4.34	[-1.33-1.39]	0.95
Implementer						
Teacher	0.92	0.55	1.69	4.43	[-0.53-2.37]	0.16
Computer	-0.27	0.49	-0.55	3.85	[-1.65-1.11]	0.61
Intervention Duration						
More than 600 minutes	-0.98	0.52	-1.87	4.37	[-2.37-0.42]	0.12
Intervention Type						
CRA	-2.05**	0.49	-4.12	3.16	[-3.59 - -0.51]	0.02
Schemas/Cognitive	-0.82*	0.38	-2.17	4.10	[-1.84-0.21]	0.09
Other	-0.35	0.22	-1.57	2.55	[-1.14-0.44]	0.23
Intervention Group						
Pair/Small Group	-0.03	0.46	-0.05	4.52	[-1.24-1.18]	0.96
Math Domain						
Multiplication/Division	-0.28	0.36	-0.77	2.74	[-1.50-0.94]	0.50
Fractions	0.19	0.73	0.27	4.21	[-1.80-2.19]	0.80
Evidence of Functional Relation						
Strong Evidence	-0.06	0.37	-0.17	6.04	[-0.98-0.84]	0.87

Note. MD = mathematics difficulties; CRA = concrete representational abstract
$*p < .10$ $**p < .05$

Intervention Characteristics

Analyses were conducted to determine whether intervention implementer, intervention duration, intervention type, intervention group size, and intervention domain functioned as moderators. Moderator results indicated that there were no significant differences in outcomes when the intervention was implemented by a researcher, teacher, or computer. Weighted averages ($\bar{X}_w$) were 0.94, 0.89, and 0.79 for researcher, teacher, and computer, respectively. Likewise, intervention duration and intervention group size did not moderate intervention effects. Weighted averages ($\bar{X}_w$) were 0.95 and 0.71 for interventions that were less than 600 minutes and interventions that were more than 600 minutes, respectively, and 0.85 and 1.02 for interventions delivered individually and interventions delivered in a small group, respectively.

Intervention type did moderate intervention outcomes, such that the effect size of mathematics interventions decreased when studies used schemas or cognitive instruction, as opposed to explicit instruction (β = -0.82, *SE* = 0.38, $p < 0.1$). Effect sizes were also significantly lower when studies used CRA as opposed to explicit instruction, however, given that the Satterthwaite degrees of freedom was less than four, this finding may not be considered reliable (Fisher & Tipton, 2015). There was no significant difference in outcomes for studies that used explicit instruction and other (e.g., extensive practice) interventions (β = -0.35, *SE* = 0.63, p = 0.51). However, it is noted that the degrees of freedom were also less than four for this comparison, and as such the finding may not be considered reliable. This finding may be a result of the low number of studies that used “other” interventions. Weighted averages ($\overline{X}_w$) were 1.14, 0.59, 1.13, and 1.12 for explicit instruction, CRA, schemas/cognitive instruction, and other interventions, respectively.

Study Quality

Analyses were conducted to determine whether study quality moderated intervention effects. Results indicated that study quality did not moderate intervention effects. There was no significant difference in outcomes for studies with strong and moderate evidence of a functional relation. Weighted averages ($\overline{X}_w$) were 0.95 and 0.89 for strong and moderate evidence, respectively

Table 5

Summary of Weighted Effect Size Averages by Moderator

Variables	*n*	$\bar{X}_w$
Grade Level		
Grade 2	12	0.67
Grade 3	9	1.11
Grade 4 to 5	10	0.92
Disability		
MD	19	0.97
LD	12	0.82
Interventionist		
Researcher	21	0.94
Teacher	7	0.89
Computer	3	0.79
Intervention Duration		
Less than 600 minutes	23	0.95
More than 600 minutes	8	0.71
Intervention Type		
Explicit	8	1.14
CRA	12	0.59
Schemas/Cognitive	7	1.13
Other	4	1.12
Intervention Group		
Individual	23	0.85
Pair/Small Group	8	1.02
Math Domain		
Addition/Subtraction	21	0.94
Multiplication/Division	7	0.79
Fractions	3	1.36
Evidence of Functional Relation		
Strong Evidence	14	0.95
Moderate Evidence	17	0.89

Note. MD = mathematics difficulties; LD = specific learning disability in mathematics; CRA = concrete representational abstract

Chapter 5

DISCUSSION

The purpose of the present study was to conduct a meta-analysis of single-case design (SCD) studies to evaluate the effectiveness of mathematics interventions for elementary grade students with mathematics difficulties (MD) and mathematics learning disabilities (LD). Moderator analyses were also conducted to examine whether intervention effectiveness varied as a function of participant characteristics (i.e., grade level and MD/LD status), intervention characteristics (i.e., intervention type, intervention domain, intervention implementer, intervention duration, and intervention group size), and study quality (i.e., evidence of a functional relation). A total of 20 studies and 31 effect sizes met eligibility criteria and were included in the meta-analysis. Studies were published during the years 2009 to 2021. In this section, I first summarize and discuss implications of the findings. Next, I address limitations of the study and propose directions for future research. Lastly, I discuss implications for practice.

What is the effectiveness of mathematics interventions for students with MD and LD in grades K-5?

Results indicated an overall significant treatment effect of mathematics interventions on students with MD and LD in grades K-5. The overall weighted effect size estimate of 0.91 indicates improvements from the baseline to treatment phase for students who received mathematics interventions. Likewise, all effect sizes were positive, indicating that on average, all students showed improved outcomes during the treatment phase (i.e. mathematics intervention). These findings are also consistent with prior meta-analyses of single-case research. However, in contrast to prior research, the present meta-analysis evaluated intervention effectiveness across mathematics domains and intervention types. For example, Lei et al. (2020) and Shin et al. (2020) both found positive effects for word problem solving interventions. Likewise, Burns et al. (2010) found positive effects of acquisition and fluency interventions for students in Grades 2-6, while Methe et al. (2012) found positive effects of interventions targeting addition and subtraction skills in students in K-6. As such, the present study provides further evidence regarding the effectiveness of mathematics interventions for students with MD and LD when using single-case research design.

Does intervention effectiveness vary as a function of participant characteristics (i.e., grade level and MD/MLD status)?

Grade Level

Results indicated that student grade level moderated intervention effects, such that larger intervention effects were found for students in Grade 3, as opposed to students in Grade 2. As previously mentioned, no studies included students in kindergarten and first grade, and as such the K-2 category only included students in Grade 2. There was no significant difference in outcomes for student in Grades 2 and Grades 4-5. However, when comparing students in Grade 2 with students in Grades 4-5, the degrees of freedom was less than four and as such this finding must be interpreted with caution. When conducting RVE analyses, the degrees of freedom are adjusted for small samples and if they are less than four, results may be considered unreliable (Fisher & Tipton, 2015). Along with the moderator analyses, weighted effect size averages were calculated for each grade level. Results indicated that on average, weighted effect sizes were larger for students in Grade 3, followed by students in Grade 4-5, and students in Grade 2.

A possible explanation for the lower effect sizes found for students in Grades 4-5 compared to students in Grade 3 is that students with disabilities tend to show slower rates of growth across the years, compared with typically achieving students, with gains decreasing as students move up a grade (Bloom et al., 2008; Wei et al., 2013). Thus, students in Grade 4-5 may need more time to demonstrate gains from interventions. By this reasoning, it would be expected that students in Grades 2 show larger intervention effects than students in Grade 3, which was not the case in the present study. It is unclear as to why students in Grade 3 showed larger intervention effects compared to students in Grade 2. A possible explanation is that 7 of the 9 effect sizes for students in Grade 3 were from studies conducted by Flores and colleagues, whereas 8 of the 12 effect sizes for students in Grade 2 were conducted by Dennis and colleagues. As such, effect size estimates may have been influenced by differences in methodological approaches. For example, Flores and colleagues measured treatment outcomes with "number correct for digit," whereas Dennis and colleagues primarily used "total problems correct."

In line with findings from the present study, Methe et al. (2012) and Lei et al. (2020) both found that grade level moderated intervention effects. Specifically, Methe et al. found that students in third grade showed the largest effect sizes followed by students in fourth and fifth grade. However, their analysis did not account for dependency in the data. Lei et al. also found significant differences in intervention effects between students in third and fourth grade, with students in fourth grade showing higher effect sizes.

However, this study only included students who were English Learners. Thus, older students may have shown larger treatment effects because they had stronger knowledge of mathematics vocabulary. In contrast to the present study, Burns et al. (2010) found no significant treatment effects across student grade level. However, unlike the present study, their sample included students with a range of disabilities, including students with learning disabilities, intellectual disability, and behavioral disabilities. In addition, the author did not conduct moderator analyses, but instead compared differences across means (e.g., Kruskal-Wallis Test) to assess for significant differences across grade level, which does not account for dependency issues or control for effects from other variables.

Disability

Findings indicated that disability status did not moderate treatment effects. Students with MD and LD both demonstrated improved outcomes following mathematics interventions and no significant differences in outcomes were found. This finding is consistent with single-case research (Peltier et al., 2020) and group design-research studies (Chodura et al., 2015; Jitendra et al., 2020; Lein et al., 2020), which have reported similar outcomes for students with MD and LD. These findings suggest that mathematics interventions have similar effects for students regardless of their level but also suggest that students with MD, who are at-risk for developing a disability, have common needs with students with learning disabilities. As such, they both benefit from mathematics interventions. For example, Peltier et al. (2020) found that students with MD and LD both showed similar treatment outcomes; however, students with behavioral disorders showed lower effect sizes. This may provide further evidence to support the notion that students with MD and LD share common needs and thus would benefit from similar intervention approaches, as opposed to students with behavioral disorders who may be better served by interventions targeting both mathematics instruction and behavior supports. On a similar note, there is a large percentage of students in the United States who struggle in mathematics without a formal diagnosis (NAEP, 2019), and as such it is important to recognize that students may benefit from intervention regardless of their disability status.

Does intervention effectiveness vary as a function of intervention characteristics (i.e., intervention implementer, intervention duration, intervention type, group size, math domain)?

Intervention Implementer

Findings indicated that there were no significant differences in treatment outcomes when the intervention was implemented by a researcher, teacher, or computer. This finding is consistent with Peltier et al. (2020) and suggests that teachers can implement interventions with similar effectiveness as researchers. However, the effectiveness of mathematics interventions delivered primarily through computer was not clear, as the present study only included 3 studies in which students received intervention through a computer program. Other research has found conflicting results, which suggest that intervention implementer does impact intervention effectiveness. For example, Lei et al. (2020) found that effect sizes were higher when interventions were delivered by teachers or by both teachers and researchers, as opposed to only researchers. However, the discrepant findings may be explained by the intervention focus and intervention type.

Lei et al. focused on delivering interventions to ELLs and as such the interventions included instruction in both mathematics and reading comprehension, which may suggest that the intervention implementer may have additional training to work with ELLs and to provide instruction in reading comprehension. Dennis et al. (2016) also found that intervention implementer moderated intervention effects; however, the authors caution that results were based on a small number of effect sizes.

Intervention Duration

Findings indicated that intervention duration did not moderate intervention effects. Intervention effects were similar when interventions were either less than 600 minutes (i.e., 10 hours) or more than 600 minutes. This finding is promising as it suggests that mathematics outcomes can be improved in students within a short period of time, and it is especially significant given many of the challenges encountering public schools, including a widespread shortage of teachers in the United States (Sutcher et al., 2019; Bryner et al., 2021). Lei et al. (2020) also found no significant differences for interventions lasting fewer than 10 hours and interventions lasting more than 10 hours. In contrast, group-design research found evidence that interventions lasting more than 10 hours are more effective (Jitendra et al., 2018; Stevens et al., 2018) than interventions lasting fewer hours. However, these findings may have been a function of the study sample, as both studies included students from middle school and high school, and as such, it is possible that older students may have more severe needs that may require more intensive intervention. Likewise, it is possible that interventions may need to last longer when involving a large group of students, which tends to be the case in group-design research. For example, Jitendra et al. (2018) included studies in which the sample size was more than 20 students. In the present meta-analyses, the largest group of students comprised only 4 students. Thus, findings suggest that in single-case research, intervention duration does not moderate intervention effects; however, it is possible that duration may play a role with larger groups of students.

Group Size

Findings indicated that intervention group size did not moderate intervention effects. Studies that delivered intervention individually or in a small group (i.e., 2 or more students) demonstrated similar positive treatment effects. This finding is consistent with prior research in that there were no significant differences between small group and individual instruction (Dennis et al., 2016; Jitendra et al. 2018; Lein et al., 2020). These findings are promising as they suggest that students requiring intensive intervention can benefit from interventions delivered both one-on-one and in a small group setting. Small intervention groups, as opposed to one-to-one intervention can have many benefits. For example, a small group may provide a setting that allows teachers to facilitate discussion among students and provides opportunities for students to practice their mathematical language, strategies which are known to help promote learning in students with MD and LD (Fuchs et al., 2021)

Intervention Type

Findings indicated that intervention type moderated intervention outcomes. All intervention types produced positive effect sizes, indicating that students with MD and LD in mathematics benefit from a range of intervention approaches. First, effect sizes were larger when interventions used explicit instruction, as opposed to schema-based instruction or cognitive strategy instruction. As noted earlier, given the small number of studies included in the present meta-analysis, schema-based instruction and cognitive instruction were collapsed into one category. Prior research has demonstrated that schema instruction and cognitive strategy instruction are both effective approaches when teaching students with MD and LD (Jitendra et al., 2013; Jitendra & Star, 2011; Fuchs et al., 2021; Swanson et al., 2013; Swanson et al., 2014). Schema-based and cognitive instruction both involve the use of heuristics that allow students to develop a deeper understanding of mathematical concepts, and help students organize information so that they can solve the problem (Fuchs et al., 2021; Montague & Dietz, 2009; Montague et al., 2011). Cognitive and metacognitive strategies, such as analyzing the problem, monitoring strategy use, and evaluating the solution, are embedded within schema-based instruction (Jitendra et al., 2013; Jitendra et al., 2016; Jitendra et al., 2015). In short, both approaches involve the use of cognitive and metacognitive processes to arrive at a solution to the problem. However, despite the similarities between schema-based instruction and cognitive strategy instruction, a recent meta-analysis found that effect sizes were lower for studies that used cognitive strategy instruction compared to schema-based instruction (Lein et al., 2020). As such, it is possible that the overall effect size for schema-based instruction in the present study would have been larger if it had not been combined into one category with cognitive strategy instruction.

Treatment effects were also higher when studies used explicit instruction as opposed to CRA; however, these findings must be interpreted with caution given the low number of degrees of freedom. Lastly, there were no significant differences when comparing explicit instruction with "Other" interventions (e.g., extensive practice, independent practice, game-based learning), which may also be explained by the low number of studies ($n = 4$) that used "Other" interventions.

Prior meta-analyses have also considered intervention type or components as a possible moderator. However, inconsistencies exist in how interventions are defined across studies. For example, Methe et al. (2012) found positive treatment effects for all intervention types, but the study had six different categories, including contingent reinforcement, cover-copy-compare, interspersal, speed-based intervention, combined, and CRA. Unlike the present study, Methe et al. (2012) included behavior-based interventions, as well as fluency interventions. Nonetheless, consistent with findings from the present study, Methe et al. also found large treatment effects for CRA interventions. However, the authors noted that only one study included CRA. Additionally, Peltier et al. (2020) found positive treatment effects for interventions that used explicit instruction when teaching lessons using manipulatives, which is consistent with the present findings suggesting that explicit instruction is an effective approach when working with students who are struggling with mathematics. The use of explicit instruction to support the needs of students with MD and LD is also supported by extensive research (Doabler et al., 2015; Fuchs et al., 2021; Gersten et al., 2009; Kroesbergen & Van Luit, 2003; Lein et al., 2020). In other words, students benefit from strategies embedded within explicit instruction, such as step-by-step instruction, teacher modeling and explaining how to solve problems, frequent opportunities for students to practice and demonstrate conceptual understanding, and continuous corrective feedback from teachers to correct errors or reduce misconceptions.

Mathematics Domain

Findings indicated that mathematics domain did not moderate intervention effects. Treatment effects were all positive but did not significantly vary across the three different mathematics domains (i.e., addition/subtraction, multiplication/division, and fractions). These findings are consistent with prior single-case and group design research which has found mathematics domain or topic does not moderate intervention effects (Jitendra et al., 2018; Lein et al., 2020; Peltier et al., 2020).

Does intervention effectiveness vary as a function of study quality?

Study quality did not moderate intervention effects. Positive treatment effects were found for studies with moderate and strong evidence of a functional relation. Based on recommendations from the WWC guidelines, effect sizes were not calculated for studies that demonstrated weak evidence of a functional relation. Likewise, prior to determining the quality of evidence of a functional relation, studies were evaluated against the WWC standards to determine whether they met standards with or without reservations. Studies that did not meet standards, were removed, and were not included in the present study. Findings from the present study are consistent with prior single-case and group design research which has found that study quality does not moderate intervention effects (Peltier et al., 2020; Shin et al., 2020; Stevens et al., 2018). While Dennis et al. (2016) found that study quality did moderate intervention effects, they evaluated quasi experimental studies and used quality indicators for group experimental and quasi-experimental research in special education.

Limitations and Future Directions

Limitations of the present meta-analyses must be considered when interpreting findings and for future research purposes. First, the sample size was relatively small and included only 20 studies and 31 effect sizes. Given the rigorous eligibility criteria that was used to determine whether a study could be included in the meta-analysis, many studies had to be excluded. For example, studies had to be excluded when the sample included students with disabilities other than LD and MD, such as Autism, Intellectual Disability, and ADHD, disabilities which often require behavior interventions in conjunction with academic interventions. However, it is important that future research consider this population of students and investigate which intervention approaches are most effective for students who are demonstrating mathematic difficulties along with behavior challenges. Also, given the small sample size, not all grade levels were represented in the study. For example, no studies included students in Grades K and 1. Moreover, the small sample size also hindered the meta-regression analysis that was conducted to examine for moderators. Various categories (e.g., intervention type, grade level) had to be collapsed for statistical purposes given the small number of studies within each level of the category. Along the same lines, the degrees of freedom for various effect sizes were low and such findings had to be interpreted with caution as they were not considered reliable. As such, a larger sample size may have allowed for more in-depth analyses of moderators.

Furthermore, many studies did not provide participant demographic information that would allow for more insight when examining the effectiveness of mathematics interventions. For example, cognitive ability of participants was only reported for 40% of participants or 9 studies, and mathematics academic achievement was only reported for 66% of participants or 13 studies. Research has shown that students with MD and LD demonstrate impairments in cognitive skills related to executive functioning, specifically working memory and attention processes (Andersson, 2010; Geary, 2011; Geary et al., 2007; Passolunghi & Siegel, 2004; Swanson et al., 2015) and as such examining cognitive ability and mathematics achievement as potential moderators may have provided valuable information. Likewise, while all studies reported the criteria that were used to identify whether a student had a learning disability in mathematics or had mathematics difficulties, there was significant variation in how students were selected for the studies. For example, studies selected participants based on teacher recommendation, performance on researcher-developed measures, performance on screeners, and performance on benchmark testing. Cut off scores and outcome measures also varied, with some studies requiring performance below the 25th percentile on a curriculum-based measure and others requiring less than 10 digits correct on a curriculum-based measure. As a result of the inconsistent reporting across the literature, it was not possible to group students based on level of need. This limitation further impacts our ability to understand how to address the needs of these students given that it is unclear how participants compare in their mathematics skills.

In line with this, many inconsistencies were observed in how studies reported mathematics achievement data, with some studies reporting standard scores, while others reported percentiles or scales scores. Furthermore, English learner status of participants was only reported in 8 studies. Examining ELL status as a potential moderator may be significant given the high number of ELLs in the United States (National Center for Education Statistics, 2018) as well as the research showing that ELL students with MD and LD may benefit from both mathematics and reading comprehension instruction (Lei et al., 2020). Despite researchers advocating for high quality reporting practices within single-case research, the reporting of participant data continues to need improvement (Horner et al., 2005; Kazdin, 2011; Kratochwill et al., 2022).

Moreover, the present study used the between-case standardized mean difference (BC-SMD) to calculate effect sizes and thus, studies could only be included if they used a reversal design (i.e., ABAB), multiple-baseline across participants, or multiple-probe design across participants. As such, studies that used a changing-criterion design (e.g., Kong & Orosco, 2016) or a multiple-baseline across groups (e.g., Dennis et al., 2015) could not be included in the meta-analysis. Future research should consider using effect sizes that would allow for these studies to be included in meta-analyses as they are likely to provide additional valuable information.

Furthermore, another limitation is that three effect size estimates had to be removed from the meta-analyses because they were outliers. It was determined that these effect sizes were significantly large due to floor and ceiling effects. As a result of removing the outliers, the sample for the present study was reduced in size, and studies that could have contributed to the findings and further inform the literature had to be excluded. Future research should consider using effect sizes that are more sensitive to floor and ceiling effects. On a similar note, an additional 3 effect sizes were removed prior to conducting the analyses given that they did not meet WWC standards (Kim et al., 2015) or had weak evidence for a functional relation (McKevett et al., 2019, Preston, 2016). While these effect sizes were not included in the analyses following the WWC Pilot Single-Case Design Standards (Kratochwill et al., 2010), many researchers choose to include all effect sizes into their analyses even if they do not meet standards or if they demonstrate weak evidence of a functional relation. By including all studies regardless of methodological quality, researchers can examine methodological quality as a possible moderator of treatment effects, or can conduct separate analyses, those that include all studies and others that do not include the studies with low methodological quality.

Next, given the significant overlap of intervention components across studies, it is unclear as to what instructional strategies or practices were driving the treatment effects. For example, most of the studies that used schema-based instruction or cognitive based instruction, also incorporated explicit instruction when providing instruction to students (Alghamdi et al., 2020; Seo & Bryant, 2012; Sharp & Dennis, 2017; Xin et al., 2020). However, these studies were not coded under explicit instruction because it was not the primary intervention approach in the study. Instead, explicit instruction was only coded for studies that primarily focused on teaching mathematical concepts with explicit instruction (Dennis, 2015; Dennis et al., 2016; Liu & Xin, 2017). Given the significant overlap across interventions, it may be appropriate for future research to, in addition to examining the intervention type, to also code for specific intervention strategies or components. For example, instead of coding an intervention as explicit instruction, interventions may be broken apart into specific components, such as, independent practice, guided practice, opportunities for students to respond, corrective feedback, and teacher modeling. A meta-analyses of group design studies by Dennis et al. (2016) took such an approach, coding studies for both instructional approaches (e.g., explicit-teacher led instruction, peer-assisted learning) and instructional components (e.g., control task difficulty, elaboration, task reduction). Future research of single-case studies that provide a closer examination of the instructional strategies or practices that may be contributing to student success is warranted.

Lastly, there is a need for future research to consider maintenance and transfer effects of treatment outcomes. In the present study, 74% of studies measured maintenance outcomes and on average, 86% of those participants showed maintenance effects in each study. However, there are inconsistencies in how long researchers wait to measure maintenance effects, and it is unclear as to how researchers determine whether students' maintained treatment effects. For example, for studies that reported maintenance outcomes, the follow-ups ranged from 1 to 6 weeks. Likewise, it is unclear as to what level of performance indicates whether treatment effects were maintained, and along those same lines various studies measured maintenance of treatment effects with only one data point (i.e., Flores, 2010; Flores & Hinton, 2019; Morin et al. 2017; Ok & Bryant, 2017). Lastly, the present meta-analysis did not examine transfer or generalizability effects. Specifically, the present study did not examine whether treatment effects generalized to standardized measures, such as standardized state tests or standardized tests of academic achievement. Given that all studies in the present meta-analyses used researcher-created measures as the outcome measures, future single-case research should also consider assessing treatment outcomes using tests that measure transfer or generalizability effects.

Implications for Practice

Results of this meta-analysis adds to our knowledge base on the efficacy of small-scale interventions using single-case research. Findings indicated that students with MD and LD showed improved outcomes when provided with one-on-one or small group intervention. Within a multi-tiered system of supports framework, schools can provide students with Tier 2 or Tier 3 interventions in individual or small group settings. Interventions delivered in small settings pose many benefits. Teachers can create small intervention groups based on students' needs. This may be especially important when working with students with heterogenous needs, such as students who struggle with mathematics but also have behavior challenges or students who are English Language Learners (Horner et al., 2005; Kazdin, 2011). Additionally, small groups allow teachers to facilitate discussion among students and can provide opportunities for students to develop their mathematical language, strategies which are known to help promote learning in students with MD and LD (Fuchs et al., 2021).

Findings indicate that schools can help improve the mathematics outcomes of students with MD and LD with the implementation of interventions that align with evidence-based approaches, including explicit instruction, CRA, schema-based instruction, and cognitive strategy instruction. As such, school districts have to identity and integrate interventions into their curriculum, as well as have in place the infrastructure required to successfully deliver these intervention services. For example, schools must be able to provide teachers with the training and resources necessary to deliver and implement interventions with high quality and fidelity. Our findings are promising as they suggest that interventions are effective whether implemented by a researcher or teacher, so it is possible for schools to train individuals to provide these interventions. Likewise, improved outcomes in students can be achieved with interventions lasting fewer than 10 hours, which is especially significant given the widespread shortage of teachers in the United States (Sutcher et al., 2019; Bryner et al., 2021).

Results indicated that elementary aged students showed improved mathematics outcomes after receiving intervention instruction. These findings underscore the importance of early intervention before students move on to the secondary grades. Research has shown that while students needs may become more severe as they move up a grade, their progress also tends to slow down (Bloom et al., 2008; Wei et al., 2013). As such, addressing the needs of students as early as possible should be a primary effort for schools. School districts should have systems in place that allow for the identification, remediation, and monitoring of student needs. For example, within an MTSS framework, school can adopt universal screening measures to identify students who are struggling, as well as progress monitoring measures to track student progress. Through universal screening and progress monitoring schools can make timely and informed decision by engaging in data-based decision making.

Conclusion

The purpose of the present study was to conduct a meta-analyses of single case research to evaluate the effectiveness of mathematics interventions for students with mathematics difficulties (MD) and mathematics learning disabilities (LD). The overall weighted effect size indicated that mathematics interventions produce positive outcomes for students with MD and LD. Additionally, moderator analyses indicated that grade level and intervention type moderated treatment effects, such that students in Grade 3 showed higher treatment effects than students in the lower elementary grades and interventions that used explicit instruction yielded the highest effect sizes compared to interventions that used schema-based instruction, cognitive strategy instruction, and CRA. Disability status, intervention duration, intervention group size, mathematics domain, intervention implementer, and study quality did not moderate intervention effects. However, findings from the moderator analyses must be interpreted with caution given the small number of studies within various categories. While findings are promising and indicate that students struggling with mathematics benefit from mathematics intervention, examining the effectiveness of mathematics interventions using meta-analytic methods is hindered by small sample sizes and limitations of BC-SMD effect sizes which require that studies with certain research designs (e.g., changing criterion) and low number of participants be excluded. As such, researchers must consider ways to maximize the number of studies that could be included in future meta-analyses.

References

Alghamdi, A., Jitendra, A. K., & Lein, A. E. (2019). Teaching students with mathematics disabilities to solve multiplication and division word problems: The role of schema-based instruction. *ZDM Mathematics Education, 52*, 123-137. https://doi.org/10.1007/s11858-019-01078-0.

Andersson, U. (2008). Mathematical Competencies in Children With Different Types of Learning Difficulties. *Journal of Educational Psychology*, *100*(1), 48–66. https://doi.org/10.1037/0022-0663.100.1.48

Andersson, U. (2010). Skill Development in Different Components of Arithmetic and Basic Cognitive Functions: Findings From a 3-Year Longitudinal Study of Children With Different Types of Learning Difficulties. *Journal of Educational Psychology*, *102*(1), 115–134. https://doi.org/10.1037/a0016838

Bailey, D. H., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science*, *17*(5), 775–785. https://doi.org/10.1111/desc.12155

Baker, D. L., Richards-Tutor, C., Sparks, A., & Canges, R. (2018). Review of single subject research examining the effectiveness of interventions for at-risk English learners. *Learning Disabilities Research & Practice: A Publication of the Division for Learning Disabilities, Council for Exceptional Children*, *33*(2), 64–74. https://doi.org/10.1111/ldrp.12160

Baker, S., Gersten, R., & Lee, D.-S. (2002). A Synthesis of Empirical Research on Teaching Mathematics to Low-Achieving Students. *The Elementary School Journal*, *103*(1), 51–73. https://doi.org/10.1086/499715

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.

Bouck, E. C., & Park, J. (2018). A Systematic Review of the Literature on Mathematics Manipulatives to Support Students with Disabilities. *Education & Treatment of Children*, *41*(1), 65–106. https://doi.org/10.1353/etc.2018.0003

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111. https://doi.org/10.1002/jrsm.12

Burns, M. K., Codding, R. S., Boice, C. H., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review*, *39*(1), 69-83.

Chodura, S., Kuhn, J.-T., & Holling, H. (2015). Interventions for Children With Mathematical Difficulties. *Zeitschrift Für Psychologie*, *223*(2), 129–144. https://doi.org/10.1027/2151- 2604/a000211

Council for Exceptional Children. (2014*). CEC standards for evidence-based practices in special education.*

Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). The handbook of research synthesis and meta-analysis (2nd ed.). Russell Sage Foundation.

Dennis, M. S. (2015). Effects of tier 2 and tier 3 mathematics interventions for second graders with mathematics difficulties. *Learning Disabilities Research & Practice*, *30*(1), 29-42.

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.

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

Dossey, J., McCrone, S. & Halvorsen, K. (2016). *Mathematics education in the United States 2016: A capsule summary book written for the thirteenth International Congress on Mathematical Education* (ICME-13). Reston, VA: National Council of Teachers of Mathematics.

Dougherty, C. (2003). Numeracy, literacy and earnings: evidence from the National Longitudinal Survey of Youth. *Economics of Education Review*, *22*(5), 511–521. https://doi.org/10.1016/S0272-7757(03)00040-2

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-463.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj*, 315(7109), 629-634.

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*.

Fisher, Z., Tipton, E., Zhipeng, H., & Fisher, M. Z. (2017). Package 'robumeta'.

Fuchs, L. S., Malone, A. S., Schumacher, R. F., Namkung, J., & Wang, A. (2017). Fraction intervention for students with mathematics difficulties: Lessons learned from five randomized controlled trials. *Journal of Learning Disabilities*, *50*(6), 631-639.

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., and 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 from http://whatworks.ed.gov/.

Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: a decade later. *Journal of Learning Disabilities*, *45*(3), 195–203. https://doi.org/10.1177/0022219412442150

Gage, N. A., Cook, B. G., & Reichow, B. (2017). Publication bias in special education meta-analyses. *Exceptional Children*, *83*(4), 428-445.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, *37*(1), 4–15. https://doi.org/10.1177/00222194040370010201

Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental and Behavioral Pediatrics: JDBP*, *32*(3), 250–263. https://doi.org/10.1097/DBP.0b013e318209edef

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, *78*(4), 1343–1359. https://doi.org/10.1111/j.1467-8624.2007.01069.x

Geary, D. C., Hoard, M. K., & Hamson, C. O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology*, *74*(3), 213-239.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical Cognition Deficits in Children With Learning Disabilities and Persistent Low Achievement: A Five-Year Prospective Study. *Journal of Educational Psychology*, *104*(1), 206–223. https://doi.org/10.1037/a0025398

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of Number Line Representations in Children With Mathematical Learning Disability. *Developmental Neuropsychology*, *33*(3), 277–299. https://doi.org/10.1080/87565640801982361

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

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics Instruction for Students With Learning Disabilities: A Meta-Analysis of Instructional Components. *Review of Educational Research*, *79*(3), 1202–1242. https://doi.org/10.3102/0034654309334431

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, *38*(4), 293-304.

Field, A., Miles, J., & Field, Z. (2012). Discovering statistics using R. Sage.

Hassinger-Das, B., Jordan, N. C., Glutting, J., Irwin, C., & Dyson, N. (2014). Domain-general mediators of the relation between kindergarten number sense and first-grade mathematics achievement. *Journal of Experimental Child Psychology*, *118*, 78–92. https://doi.org/10.1016/j.jecp.2013.09.008

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*(3), 224-239.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, *4*(4), 324-341.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *Bmj*, *327*(7414), 557-560.

Hinton, V. M., & Flores, M. M. (2019). The Effects of the Concrete-Representational-Abstract Sequence for Students at Risk for Mathematics Failure. *Journal of Behavioral Education*, *28*(4), 493–516. https://doi.org/10.1007/s10864-018-09316-3

Horner, R. H., Carr, E. G., Halle, J., McGee, G., & 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/10.1177/001440290507100203

Hughes, E.M., Witzel, B. S., Riccomini, P. J., Fries, K. M., & Kanyongo, G. Y. (2014). A Meta-Analysis of Algebra Interventions for Learners with Disabilities and Struggling Learners. *The Journal of the International Association of Special Education*, *15*(1), 36.

Jayanthi, M., Gersten, R., Baker, S. (2008). Mathematics instruction for students with learning disabilities or difficulty learning mathematics: A guide for teachers. Portsmouth, NH:RMC Research Corporation, Center on Instruction.

Jitendra, A. K., Alghamdi, A., Edmunds, R., McKevett, N. M., Mouanoutoua, J., & Roesslein, R. (2020). The Effects of Tier 2 Mathematics Interventions for Students With Mathematics Difficulties: A Meta-Analysis. *Exceptional Children*, *87*(3), 307–325. https://doi.org/10.1177/0014402920969187

Jitendra, A. K., Dupuis, D. N., Rodriguez, M. C., Zaslofsky, A. F., Slater, S., Cozine-Corroy, K., & Church, C. (2013). A randomized controlled trial of the impact of schema-based instruction on mathematical outcomes for third-grade students with mathematics difficulties. *The Elementary School Journal*, *114*(2), 252-276.

Jitendra, A. K., Nelson, G., Pulles, S. M., Kiss, A. J., & Houseworth, J. (2016). Is mathematical representation of problems an evidence-based strategy for students with mathematics difficulties?. *Exceptional Children*, *83*(1), 8-25.

Jitendra, A. K., Petersen-Brown, S., Lein, A. E., Zaslofsky, A. F., Kunkel, A. K., Jung, P. G., & Egan, A. M. (2015). Teaching mathematical word problem solving: The quality of evidence for strategy instruction priming the problem structure. *Journal of Learning Disabilities*, *48*(1), 51-72.

Jordan, N. C. (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., Resnick, I., Rodrigues, J., Hansen, N., & Dyson, N. (2017). Delaware Longitudinal Study of Fraction Learning: Implications for Helping Children With Mathematics Difficulties. *Journal of Learning Disabilities*, *50*(6), 621–630. https://doi.org/10.1177/0022219416662033

Kazdin, A. E. (2011).S*ingle-case research designs: Methods for clinical and applied settings*.Oxford University Press

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M &Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, 34(1), 26–38. https://doi.org/10.1177/0741932512452794

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.

Komsta L. outliers: Tests for outliers [R package]. Version 0.14. 2011. Retrieved May 25, 2021 from: http://CRAN.R-project.org/package=outliers

Lee, J. (2012). College for all: Gaps between desirable and actual P–12 math achievement trajectories for college readiness. *Educational Researcher*, *41*(2), 43-55

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.

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/10.1037/edu0000453

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. SAGE publications, Inc.

Mazzocco, M. M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research & Practice*, *20*(3), 142-155.

Methe, S. A., Kilgus, S. P., Neiman, C., & Chris Riley-Tillman, T. (2012). Meta-Analysis of Interventions for Basic Mathematics Computation in Single-case Research. *Journal of Behavioral Education*, *21*, 230–253. https://doi.org/10.1007/s10864-012-9161-1

Misquitta, R. (2011). A review of the literature: Fraction instruction for struggling learners in mathematics. *Learning Disabilities Research & Practice*, *26*(2), 109-119

Morgan, P. L., Farkas, G., & Qiong Wu. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, *42*(4), 306–321.https://doi.org/10.1177/0022219408331037

Murphy, M. M., Mazzocco, M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (LD) vary as a function of the cutoff criterion used to define LD. *Journal of Learning Disabilities*, *40*(5), 458-478.

National Center for Education Statistics. (n.d.). NAEP data explorer. Institute of Education Sciences, U.S. Department of Education. http://nces.ed.gov/nationsreportcard/naepdata

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standard in Mathematics*.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. U.S. Department of Education. http://www.ed.gov/mathpanel

Nelson, G., & McMaster, K. L. (2018). The effects of early numeracy interventions for students in preschool and early elementary: A meta-analysis. *Journal of Educational Psychology*, *111*(6), 1001–1022. https://doi.org/10.1037/edu0000334

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/10.1177/0022219417714773

No Child Left Behind Act of 2001, 20 U.S.C. § 6319 (2002).

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single- case research. *Exceptional Children*, *75*(2), 135-150.

Parsons, S., & Bynner, J. (1997). Numeracy and employment. *Education and Training*, *39*(2). https://doi.org/10.1108/00400919710164125

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, *88*(4), 348–367. https://doi.org/10.1016/j.jecp.2004.04.002

Peltier, C., Morin, K. L., Bouck, E. C., Lingo, M. E., Pulos, J. M., Scheffler, F. A., Suk, A., Mathews, L. A., Sinclair, T. E., & Deardorff, M. E. (2020). A Meta-Analysis of Single-Case Research Using Mathematics Manipulatives With Students At Risk or Identified With a Disability. *The Journal of Special Education*, *54*(1), 3–15. https://doi.org/10.1177/0022466919844516

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, *108*(4), 455–473. https://doi.org/10.1037/edu0000079

Powell, S. R. (2011). Solving word problems using schemas: A review of the literature. *Learning Disabilities Research & Practice*, *26*(2), 94-108.

Powell, S. R., Doabler, C. T., Akinola, O. A., Therrien, W. J., Maddox, S. A., & Hess, K. E. (2019). 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

Price, G. R., & Ansari, D. (2013). Dyscalculia: Characteristics, causes, and treatments. *Numeracy, 6*(1), 1-16.

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

Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-Comparable Effect Sizes in Multiple Baseline Designs: A General Modeling Framework. *Journal of Educational and Behavioral Statistics*. https://doi.org/10.3102/1076998614547577

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, 93, 346-362. https://doi.org/10.1037//0022-0663.93.2.346

Rivera-Batiz, F. L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources*, 313-328.

Rohatgi, A. (2020). WebPlotDigitizer (Version 4.4) [Computer software]. Retrieved from http://arohatgi.info/WebPlotDigitizer

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, 86, 638–641. doi:10.1037/0033-2909.86.3.638

Satsangi, R., Bouck, E. C., Taber-Doughty, T., Bofferding, L., & Roberts, C. A. (2016). Comparing the Effectiveness of Virtual and Concrete Manipulatives to Teach Algebra toSecondary Students With Learning Disabilities. *Learning Disability Quarterly*, *39*(4), 240–253. https://doi.org/10.1177/0731948716649754

Shadish, W. R., Hedges, L. V., Horner, R. H., & Odom, S. L. (2015). The Role of Between-Case Effect Size in Conducting, Interpreting, and Summarizing Single-Case Research. NCER 2015-002. *National Center for Education Research*.

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

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, *2*(3), 188-196.

Shadish, W. R., Zelinsky, N. A., Vevea, J. L., Kratochwill, T. R. (2016). A survey of publication practices of single-case design researchers when treatments have small or large effects. *Journal of Applied Behavior Analysis*, 49, 656–673. doi:10.1002/jaba.308

Sharp, E., & Shih Dennis, M. (2017). Model drawing strategy for fraction word problem solving of fourth-grade students with learning disabilities. *Remedial and Special Education*, *38*(3), 181-192.

Shin, M., & Bryant, D. P. (2015). A synthesis of mathematical and cognitive performances of students with mathematics learning disabilities. *Journal of Learning Disabilities*, *48*(1), 96–112. https://doi.org/10.1177/0022219413508324

Shin, M., & Bryant, D. P. (2015). A synthesis of mathematical and cognitive performances of students with mathematics learning disabilities. *Journal of Learning Disabilities*, *48*(1), 96-112.

Shin, M., Bryant, D. P., Powell, S. R., Jung, P.-G., Ok, M. W., & Hou, F. (2020). A Meta-Analysis of Single-Case Research on Word-Problem Instruction for Students With Learning Disabilities. *Remedial and Special Education: RASE*, 0741932520964918. https://doi.org/10.1177/0741932520964918

Siegler, R., Carpenter, T., Fennell, F., Geary, D., Lewis, J., Okamoto, Y., Thompson, L., & Wray, J. (2010). *Developing effective fractions instruction for kindergarten through 8th grade: A practice guide* (NCEE #2010-4039). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from whatworks.ed.gov/ publications/practiceguides.

Star, J. R., Caronongan, P., Foegen, A., Furgeson, J., Keating, B., Larson, M. R., Lyskawa, J., McCallum, W. G., Porath, J., & Zbiek, R. M. (2015). *Teaching strategies for improving algebra knowledge in middle and high school students* (NCEE 2014-4333). Washington, DC: National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. Retrieved from the NCEE website: http://whatworks.ed.gov.

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: RASE*, *39*(6), 327–340. https://doi.org/10.1177/0741932517731887

Swanson, H. L., Lussier, C., & Orosco, M. (2013). Effects of cognitive strategy interventions and cognitive moderators on word problem solving in children at risk for problem solving difficulties. *Learning Disabilities Research & Practice*, *28*(4), 170-183.

Swanson, H. L., Orosco, M. J., & Lussier, C. M. (2014). The effects of mathematics strategy instruction for children with serious problem-solving difficulties. *Exceptional Children*, *80*(2), 149-168.

Swanson, H. L., Lussier, C. M., & Orosco, M. J. (2015). Cognitive strategies, working memory, and growth in word problem solving in children with math difficulties. *Journal of Learning Disabilities*, *48*(4), 339-358.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: both executive and phonological processes are important. *Journal of Experimental Child Psychology*, *79*(3), 294–321. https://doi.org/10.1006/jecp.2000.2587

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: practical considerations including a software tutorial in Stata and spss. *Research Synthesis Methods*, *5*(1), 13–30. https://doi.org/10.1002/jrsm.1091

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling Complex Meta-analytic Data Structures Using Robust Variance Estimates: a Tutorial in R. *Journal of Developmental and Life-Course Criminology*, *2*(1), 85–112. https://doi.org/10.1007/s40865-016-0026-5

Valentine, J.C., Tanner-Smith, E.E., Pustejovsky, J.E., & Lau, T.S. Between-case standardized mean difference effect sizes for single-case designs: A primer and tutorial using the scdhlm web application. Oslo, Norway: The Campbell Collaboration. Retrieved from: www.campbellcollaboration.org/ doi: 10.4073/cmdp.2016.1

Watt, S. J., Watkins, J. R., & Abbitt, J. (2016). Teaching Algebra to Students With Learning Disabilities: Where Have We Come and Where Should We Go? *Journal of Learning Disabilities*, *49*(4), 437–447. https://doi.org/10.1177/0022219414564220

Wei, X., Lenz, K. B., & Blackorby, J. (2013). Math growth trajectories of students with disabilities: Disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education*, *34*(3), 154-165.

What Works Clearinghouse. (2020). What Works Clearinghouse Standards Handbook, Version4.1. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. This report is available on the What Works Clearinghouse website at https://ies.ed.gov/ncee/wwc/handbooks.

Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K. R., & Ogbuehi, P. (2012). *Improving mathematical problem solving in grades 4 through 8: A practice guide* (NCEE 2012-4055). Washington, DC: National Center for EducationEvaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch/.

Xin, Y. P., & Jitendra, A. K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *The Journal of Special Education*, *32*(4), 207-225.

Xin, Y. P., & Zhang, D. (2009). Exploring a conceptual model-based approach to teaching situated word problems. *The Journal of Educational Research*, *102*(6), 427-442.

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

Appendix A

Author:
Title:
Journal:

STUDY FEATURES		
DEMOGRAPHICS	**Categories**	**Page #**
1. Number of cases (*report total # of cases and # that are LD or MD)*:	LD: MD:	
2. Subject selection *(how were students selected for the study?)*:	**1** = At or below cut scores on a math screening measure or benchmark assessments **2** = Teacher nomination **3** = Combined **4** = Other	
3. LD identification criteria *(e.g., district criteria for LD; IQ-achievement discrepancy)*:	**1** = District criteria for LD **2** = At or below cut scores on a math screening measure or benchmark assessments **3** = Combined	
4. Grade *(report grade level of each subject)*:	**1** = K–2 **2** = 3–4 **3** = 5 **4** = Mixed	
5. Gender *(report gender of each subject)*:	**0** = Male **1** = Female	
6. Subject demographic information (*e.g., race/ethnicity, ELL status, and eligibility for FRL; IQ and Ach. scores)*:		
7a. Race/Ethnicity White: AA/Black: Hispanic: Asian: Other: Not reported: *(Indicate the # in each category)*		
7b. ELL Status **0** = No **1** = Yes **2** = Not reported *(If yes, indicate how many)*		
7c. Eligibility for FRL **0** = No **1** = Yes **2** = Not reported *(If yes, indicate how many)*		
7d. IQ Scores **0** = Not reported **1** = Yes *(If yes, report scores for each subject)*		
7e. Achievement Scores **0** = Not reported **1** = Yes *(If yes, report scores for each subject)*		
8. Location:	**1** = Pub. school **2** = Priv. school	

INTERVENTION	**Categories**	**Page #**
9. Instruction/Intervention Setting ***(report where instruction occurred–resource class, sped class, gen ed class, other)*****:**	**1** = Gen. ed **2** = Resource room/sped class **3** = Other (library, computer lab) **4** = Not reported	
10. Interventionist ***(e.g., researcher, teacher, paraprofessional)*****:**	**1** = Researcher **2** = Teacher **3** = Teacher and Researcher **3** = other **4** = Not reported	
11. Instructional arrangement:	**1** = Individual **2** = Pairs **3** = Group *(indicate group size)*	
12. Instructional time ***(in min):*** **Total time:**	**1** = $\geq$ 600 min **2** = < 600 min	
13. Independent variable ***(describe the intervention)*****:**	**1** = Explicit/direct instruction **2** = SBI **3** = CRA/SIM **4** = Cog./metacog. strategy instruction **5** = Other	
OUTCOMES	**Categories**	**Page #**
14. Dependent variable ***(describe the measures)*****:**	**1** = *Counting and Cardinality* **2** = *Number and Operations in Base Ten (NBT)* **3** = *Operations and Algebraic Thinking (OA)* (word problem solving test) **4** = *Fractions* *(Specify the DV)*	
15. Outcome Measure:	**1** = Number of digits correct **2** = Number of problems correct **3** = % correct **4** = Meant total score	
16. Fidelity of implementation/Treatment Integrity *(%)***:**	**0** = No **1** = Yes *(If yes, report mean scores)*	
17. Inter-assessor agreement *(report for each phase, if available)***:**	**0** = No **1** = Yes *(If yes, report IRA for each phase)*	
18. Social validity *(report yes or no)***:**	**0** = No **1** = Yes	

Appendix B

DESIGN STANDARDS: SINGLE-CASE (WWC)		
Research Design: *For multiple probe across student groups, there must be 3 groups of 2 subjects per group (indicate the number of groups and number of subjects).*	**1** = multiple baseline across subjects **2** = multiple baseline across groups **3** = multiple probe across subjects **4** = multiple probe across student groups **5** = multiple probe across behaviors **6** = alternating treatments with baseline across subjs. **7** = other	
Number of participants:		
WWC DESIGN STANDARDS		**Page #**
Step 1: Evidence for meeting design standards 1. IV systematically manipulated (i.e., implemented and/or withdrawn), with the researchers determining when and how changes in IV conditions bring about changes in the DV (I.e., outcome variable)	**0** = No **1** = Yes	
2a. Each DV measured systematically over time by more than one assessor.	**0** = No **1** = Yes	
2b. Inter-assessor agreement was collected in each phase and on at least 20% of the sessions in each phase (e.g., baseline, intervention). See Item 17.	**0** = No **1** = Yes *(Note the percentage)*	
2c. Inter-assessor agreement met threshold (.80 to .90 if measured by percentage agreement; 0.60 if measured by Cohen's kappa).	**0** = No 1 = Yes *(Note the IAA)*	
3. Study includes at least 3 attempts to demonstrate an intervention effect at three different points in time **OR** with three different phase repetitions (i.e., tiers within MBDs)	**0** = No **1** = Yes *(Note the number of attempts)*	
4. Phase criteria: Number of phases across participants and number of data points per phase: Multiple probe designs (MPD) must meet additional criteria. (a) Initial preintervention sessions must overlap vertically. Within the first three sessions, the design must include three consecutive probe points for each case to *Meet Pilot SCD Standards without Reservations* and at least one probe point for each case to *Meet Pilot SCD Standards with Reservations*. (b) Probe points must be available just prior to introducing the independent variable. Within the three sessions just prior to introducing the independent variable, the design must include three consecutive probe points for each case to *Meet Pilot SCD Standards without Reservations* and at least one probe point for each case to *Meet Pilot SCD Standards with Reservations*. (c) *Each case not receiving the intervention must have a probe point in a session where another case either (a) first receives the intervention or (b) reaches the prespecified intervention criterion. This point must be consistent in level and trend with the case's previous baseline points.	**2** = Met *(≥ 4 phases with ≥ 5 points in ABAB, ≥ 6 phases with ≥ 5 points in MBD, 3 consecutive probe points within the first three preintervention sessions <u>and</u> 3 probe points prior to introducing the IV in MPD, or ≥ 5 points per condition with ≤ 2 points in AT)* **1** = Partially met *(≥ 4 phases with 3–4 points in ABAB, ≥ 6 phases with 3–4 points in MBD, at least 1 probe point for each case within the first three preintervention sessions <u>and</u> at least 1 probe point prior to introducing the IV in MPD, or ≥ 4 points per condition with ≤ 2 points in AT)* **0** = Did not meet *(i.e., ≤ 3 phases or ≤ 2 points in ABAB, ≤ 5 phases or ≤ 2 points in MBD, did not include a probe point for each case within the first three preintervention session or a probe point prior to introducing the IV in MPD or ≤ 3 points per condition with ≤ 2 points in AT)*	
5.Final determination of design standards:	**2** = Met design standards without reservations **1** = Met design standards with reservation **0** = Did not meet design standards	
Only for studies that "Met Design Standards without Reservations" or "Met Design Standards with Reservation"		
Step 2: Evidence for demonstrating functional relation (a) 3 demonstrations of the intervention effect (b) Each intervention effect occurring at different points in time (c) Consistency of level, trend, and variability within a phase (d) Immediacy of the effect, proportion of data overlap between phases, consistency of data across phases (e) Any external factors and anomalies such as a sudden change of level within a phase.	**a. 0** = No **1** = Yes **b. 0** = No **1** = Yes **c. 0** = No **1** = Yes **d. 0** = No **1** = Yes **e. 0** = No **1** = Yes	

DESIGN STANDARDS: SINGLE-CASE (WWC)		
Final determination of functional relation:	**2** = Strong evidence **1** = Moderate evidence **0** = No evidence	
Only for studies deemed to have either "Strong Evidence" or "Moderate Evidence" of functional relation		
Step 3: Calculate BC-ES	Participant/group 1: Participant/group 2: Participant/group 3: Participant/group 4: Participant/group 5: Participant/group 6:	