Self-Regulated Mathematic Problem-Solving: A Meta-Analysis of Middle School Interventions for Students with Disabilities

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Emily C. Bumpus, MA

Graduate Program in Educational Studies

The Ohio State University

2022

Dissertation Committee:

Laurice Marie Joseph, Advisor

Scott L. Graves

Moira R. Konrad

Abstract

This study sought to examine the effectiveness of math problem solving interventions with self-regulated learning (SRL) components on the math problem-solving performance of middle school students with disabilities. A meta-analysis was conducted of all single-case design studies meeting inclusion/exclusion criteria between 2000-2021, resulting in a sample size of 22 studies. Inclusion criteria involved the following: a mathematic problem-solving intervention, math problem-solving performance on real-world word problems as a dependent measure, students in grades 6th-8th or aged 11-14, disability qualification information for participating students, and an observable component of SRL. Independent raters provided secondary coding for all aspects of the initial systematic search and for other components of data analysis.

Disabilities represented included intellectual disabilities, learning disabilities, Autism, ADHD/Other Health Impairment, and an auditory processing disorder. Studies took place in a variety of settings both geographically and within school buildings. Interventions involved six general categories of approaches including schema-based instruction, modified schema-based instruction, concrete-representation-abstract method variations, diagram variations, general cognitive strategy instructions, and technology-based approaches. All studies included use of a cognitive or metacognitive strategy that related to self-regulated learning.

The results of this study reported an overall large effect size (Tau-U = 0.93) for math problem-solving interventions with self-regulated learning components on math problem-solving performance for middle school students with disabilities. These results were supported through a variety of sensitivity checks. The moderator analysis did not reveal any significant differences across disability category, interventionist, group setting, intervention type, or problem type. The risk of bias assessment determined that few overall domains of bias and few studies had high risks of bias.

Future studies conducted in this area should incorporate a more targeted approach to measuring the mediating effects of SRL components on math problem-solving intervention effectiveness. Additional intervention approaches should be attempted with this population that have not been tried previously. The use of a different effect size measure is also recommended when conducting single-case design meta-analysis research.

Keywords: meta-analysis, self-regulated learning, math problem solving, middle school, disabilities

Dedication

This dissertation is dedicated to my husband, John. The amount of love and support he has given since the beginning of this journey is immeasurable. He constantly grounded me amid anxiety and self-doubt, reminding me to see my worth through the eyes of our Creator. His dedication to being a good father and a diligent worker in all ways inspired me every day to see this through to the end.

Acknowledgments

Many thanks are due to those who supported the completion of this dissertation. First and foremost, I would like to express my deepest thanks to my committee members for their support: Dr. Laurice Joseph, Dr. Scott Graves, and Dr. Moira Konrad. Additional thanks to Dr. Joseph for providing constructive feedback, kind words, and belief in my ideas related to research and practice and to Dr. Graves for offering practical suggestions, helpful advice, and unwavering belief in my abilities to meet my professional goals.

Particularly helpful to me during this time were Dr. Susan Mauck and Joshua Peri, the Quantitative Research Consultants within the College of Education and Human Ecology's Quantitative Methodology Center. Their service as a sounding board and subsequent guidance strengthened the method and results of this study immensely.

Special thanks are due to my parents, parents-in-law, family, and friends that provided countless hours of childcare, moral support, and meals. It was a relief to know that even if I had to work, my son was surrounded by his favorite people that give him so much love and joy.

I am also deeply grateful to my doctoral cohort member, Kayla Lee. We have walked through many professional and personal milestones together, and her friendship made each part of this process feel less isolating and more attainable.

Finally, I am especially thankful to my God who has sustained me and carried me through all that life has entailed – especially these past two years.

Vita

2009–2013..BSEd in Mild-Moderate Educational Needs, Ohio University

2013–2017..Intervention Specialist, Columbus Collegiate Academy – Main Street, Columbus, Ohio

2017–2018..MA in Educational Studies, The Ohio State University

2017–2019..Graduate Research Associate, Department of Educational Studies, The Ohio State University

2019–2021..Graduate Teaching Associate, Department of Educational Studies, The Ohio State University

2021–2022..Intern School Psychologist, Columbus City Schools, Columbus, Ohio

Publications

Lee, K. B., Graves, S. L., & **Bumpus, E. C.** (2022). Competency standards in cognitive assessment course assignments: A national analysis of school psychology syllabi. *Training and Education in Professional Psychology*. Advance online publication. https://doi.org/10.1037/tep0000409

Bumpus, E. C., Vinco, M. H., Lee, K. B., Accurso, J. F., & Graves, S. L. (2020). The consistency of expectations: An analysis of learning objectives within cognitive assessment course syllabi. *Teaching of Psychology,* 1–7. https://doi.org/10.1177/0098628320965248

Cannella-Malone, H. I., **Bumpus, E. C.,** & Sun, X. (2020). Using a job-matching assessment to inform skills to target with video prompting. *Advances in Neurodevelopmental Disorders*, *4*, 471–479. https://doi.org/10.1007/s41252-020-00182-7

Miller, L. T., **Bumpus, E. C.,** & Graves, S. L. (2020). The state of cognitive assessment training in school psychology: An analysis of syllabi. *Contemporary School Psychology*, *25*, 149–156. https://doi.org/10.1007/s40688-020-00305-w

Cannella-Malone, H. I., Schmidt, E. K., & **Bumpus, E. C.** (2018). Assessing preferences using eye gaze technology for individuals with significant intellectual and physical disabilities. *Advances in Neurodevelopmental Disorders, 2*(3), 300–309. http://doi.org/10.1007/s41252-018-0072-6

Fields of Study

Major Field: Educational Studies, School Psychology

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

The Individuals with Disabilities Education Act (IDEA) of 1975 sought to provide an appropriate education for all children and students with a disability served within a public school that received federal funding. The history of intervention research started well before the implementation of IDEA, but researchers continue to study and discover effective teaching practices for this diverse population of students based on varying cognitive, physical, behavioral, or academic needs across academic contexts (Scammacca et al., 2016). Today there is a growing movement to increase research and the use of evidence-based practices in classrooms for mathematics specifically (VanDerHeyden, 2021; The Science of Math, n. d.). Mathematics is a multi-faceted subject, most often split into pure mathematics – calculations – and applied mathematics – real-world word problems involving a myriad of deep conceptual understandings and skills beyond calculations alone. As Trena Wilkerson (2022), the president of the National Council of Teachers of Mathematics, recently stated in her President's Message, "building and developing mathematical processes and practices through a coherent, cohesive approach to mathematical concepts and procedures that values students' positive mathematical identity is essential to developing this deep understanding" (para. 3). To that end, more research has been published on mathematic problem-solving interventions and their effectiveness with students with disabilities in recent years. Of particular interest is the long-standing connection between mathematic reasoning and self-regulated learning (Meverach et al., 2018).

Self-Regulated Learning

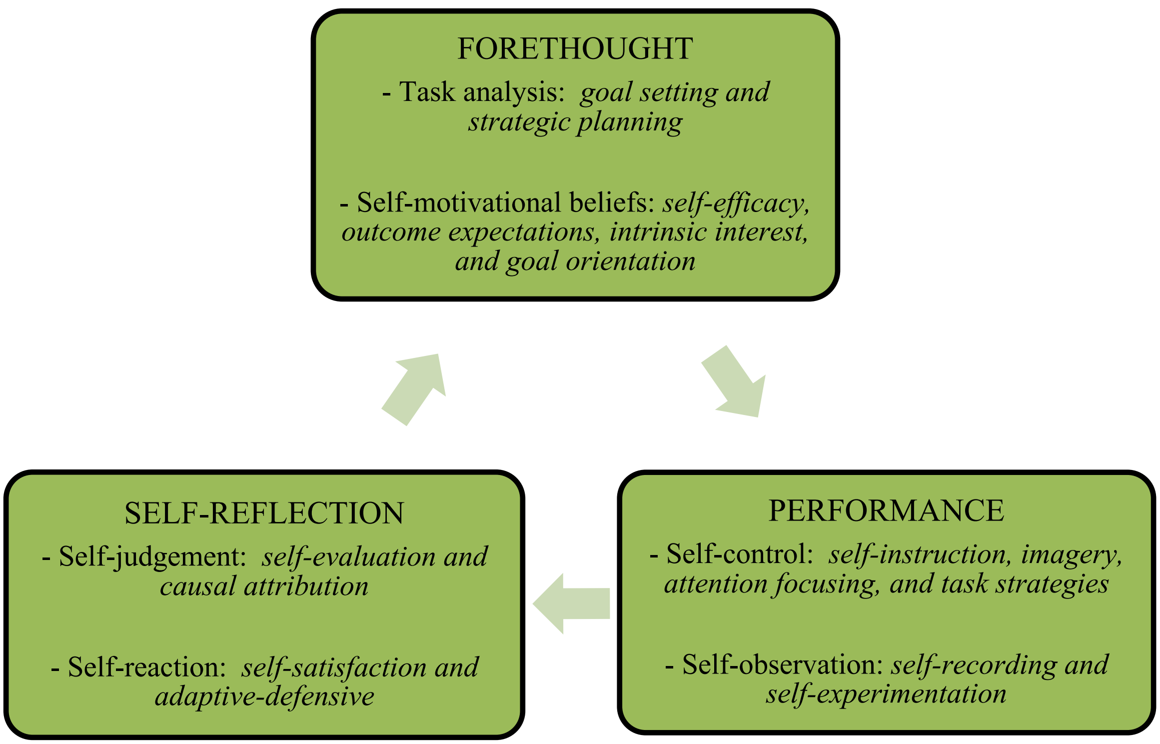
The field of research around self-regulation and self-regulated learning (SRL) is dynamic and complex with a history of multi-faceted approaches to understanding how SRL impacts individual performance (Schunk & Greene, 2018). Schunk and Greene broadly define self-regulation as "the ways that learners systematically activate and sustain their cognitions, motivations, behaviors, and affects toward the attainment of their goals" (p.1). Within this definition, there are multiple broad theoretical approaches to SRL that adjust this definition based on the different approach. The social cognitive approach, as presented by Bandura and Zimmerman, posits that the systematic activation and sustaining of cognition is encompassed and influenced by outward behaviors, environmental factors, and personal beliefs intertwined (Usher & Schunk, 2018). Through a social cognitive approach, individuals implement self-regulation strategies through imitation from models observed in their environment (Hoyle & Dent, 2018). Cognitive theoretical approaches, on the other hand, focus more on the innate "coming to know" progression of cognition related to how individuals regulate or control both taking in and producing information than on external social influences of the process (Winne, 2018, p. 36). This includes another area of theoretical approaches specifically analyzing how individuals engage in metacognition (an awareness and analysis of one's own thought process) and how these processes interplay with the cognitive process itself (Winne, 2018). Another approach to SRL, the information-processing perspective, postulates that SRL develops most through repeated practice and allowance for tweaking of metacognitive strategies used to increase efficiency during subsequent attempts (Hoyle & Dent, 2018; Winne, 2018). This brings on a dynamic model of development as SRL abilities may be learned and practiced both in a global context as an individual experiences the world and then within more specific academic learning contexts for targeted skill development including temperament, executive attention, executive functions, delay of gratification, verbalization, associative memory, and metamemory (Hoyle & Dent, 2018). Yet another approach and definition of SRL focuses on the roles of motivation and affect within the cognitive process. Research with this focus describes SRL "as a series of events that ensure goal-directed, deliberate regulation of processing in learning tasks" (Efklides et al., 2018). Specifically, Efklides et al. postulated that not only do motivation, affect, and metacognition each have separate impacts on regulation, but the interactions between each of these components play a role in how well an individual may complete SRL. This theory includes the mediator of emotion and frustration within a learning process over one's ability to control and choose future cognitive steps (Efklides et al., 2018).

Students who struggle with SRL have been shown to have poor academic achievement and behavior problems, as well as issues with physical health such as diet, medicines, and chronic disease (Zimmerman, 2000). Self-regulation additionally plays a critical role in an individual's ability to engage with higher-order metacognitive skills in academic settings such as planning, monitoring, control, and reflection (Mevarech et al., 2018). These metacognitive skills tie into Zimmerman's foundational cyclical feedback phases: forethought, performance, and self-reflection. Each phase has categories and multiple subprocesses respectively that detail the self-regulatory cycle (see Figure 1). The feedback an individual receives in any of these phases impacts adjustments for future efforts at the same task or in the same phase (Zimmerman, 2000).

It is important for students to learn cognitive self-regulation because it can improve positive behaviors, academic achievement, and self-view all to in turn create life-long independent learners (Hoyle & Dent, 2018). The theoretical information processing approach to SRL reinforces the importance of practice to learning SRL, particularly during the secondary years (Hoyle & Dent, 2018). While the information processing theory has roots in Zimmerman's (2000) social cognitive theory, it expands more on subprocesses within the cyclical phases and shifts focus to the exact cognitive and metacognitive actions occurring during learning and self-regulation. With a desire to focus on students' self-regulation in classroom learning

Figure 1

Zimmerman's (2000) Cyclical Phases, Subprocesses, and Forms of Self-Regulation



Note. Adapted from "Attaining self-regulation: A social cognitive perspective," by B. Zimmerman, (2000), in M. Boekaerts, P. Pintrich, and M. Zeidner's (Eds.) Handbook of Self-Regulation, p. 16. Copyright 2000 by Academic Press. Reprinted with permission of Elsevier Science & Technology Books; permission conveyed through Copyright Clearance Center, Inc.

environments, it is critical to recognize the influence of social and environmental aspects on students' ability to engage in self-regulatory learning (Butler & Cartier, 2018). Of specific importance, the intentional design – or lack thereof – of learning activities, appropriate social influences, and the provision of modifiable self-regulation support within classroom instruction and intervention are key environmental influences, especially for students with disabilities (Butler & Cartier, 2018; Mason & Reid, 2018).

Interventions and classroom practices that have sought to target self-regulation skills have been studied previously through a variety of approaches within the classroom setting (Axe et al., 2019; Fuentes, 2019; Kornick & deFur, 2016; Ness & Middleton, 2012). Ness and Middleton (2012) detail different methods within the classroom based on the planning, performance, or self-evaluation phase of SRL desired to be targeted: organizational/motivational checklists for the planning phase, mnemonics detailing steps of a given cognitive strategy for the performance phase, and dedicated time to structured graphs and discussion related to performance for the self-evaluation phase. General guidelines for ways teachers can embed self-regulation practice into their general education classrooms include classroom organization, clear establishment of expectations/rules/routines, using checklists, giving students choice and a voice within the classroom, modeling SRL language and skills, setting goals, using strategic questions, and providing positive feedback (Korinek & deFur, 2016). Oftentimes implementing SRL interventions related to academic achievement also includes looking at students' self-efficacy levels in that area (Fuentes, 2019; Marshall et al., 2019; Cuenca-Carlino et al., 2016; Choi & Walters, 2018).

Furthermore, there has been a recent increase in interest to closer analyze the ways in which components of self-regulation are measured in real-time application (Schunk & Greene, 2018). These methodologies include self-report questionnaires (Wolters & Won, 2018), think-aloud protocols (Greene et al., 2018), microanalytic methods (Cleary & Callan, 2018), in-depth case studies (Butler & Cartier, 2018), trace data (Bernacki, 2018), and data mining (Biswas et al., 2018). New approaches are attempting to measure the more dynamic nature of SRL compared to previous reliance on self-reports from individuals that were not as reliable (Schunk & Greene, 2018). Research in interventions targeting SRL within academic contexts has been successful but could improve by capturing how a student adjusts their cognitive methods during the task itself (Schunk & Greene, 2018).

Mathematic Problem Solving

A recent push in the realm of mathematics research, called the Science of Math, focuses on increasing the use of evidence-based practices in classrooms and calling attention to harmful misconceptions that are prevalent within many mathematics classrooms across the United States today (VanDerHeyden et al., 2021). The Science of Math movement advocates for the use of successfully research-based math strategies for learners with disabilities to be implemented in class-wide instruction as a gold standard for general math instruction regarding computation and problem-solving (VanDerHeyden et al., 2021). Specifically, the use of explicit instruction is a valued approach for teaching mathematics effectively to students in K-12 school settings (The Science of Math, n.d.). VanDerHeyden et al. (2021) outline additional specific evidence-based practices that are effective within mathematics classrooms. The research base supports these practices as they tie directly to specific components of how effective mathematic problem-solving occurs for students.

Academic skills in mathematics build upon one another, starting with straight-forward computations, such as addition, and building up to more complex tasks such as calculus.

Teachers are directed through best practices to utilize problem solving as a core process through which to teach key mathematic concepts (National Council of Teachers of Mathematics [NCTM], 2000; Van de Walle, 2007). Mathematic problem solving is a task in which one engages without knowing an effective solution at the outset (NCTM, 2000). More specifically, NCTM (2000) defines problem solving as ensuring students have skills to do the following: (1) learn new knowledge through problem solving skills; (2) generalize these skills into other contexts; (3) adapt strategies based on the given situation; and (4) reflect on the ongoing process as it unfolds. As such there are many necessary skills within the problem-solving process that are important to understand.

The core processes of problem solving are two-fold: problem execution and problem representation (Montague, 2006). Students must be adept at both components in order to successfully complete mathematical problem solving; the problem must be represented accurately in order to create an opportunity for accurate computational execution (Babakhani, 2011). Furthermore, the problem-solving process is fluid, where it may be necessary for the individual to move back and forth between these two core processes multiple times before discovering an appropriate solution (Kikas et al., 2020).

As might be easily deduced, problem execution relies on accurate computation fluency and arithmetic (Tolar et al., 2016). Instruction in computation and fluency is essential in order to be successful with higher-order math skills – and there is a growing call for research-based practices in the classroom regarding how students are taught these skills specifically (VanDerHeyden et al., 2021). Yet other studies have also shown that explicit instruction in word-problem strategies – mathematic problem solving – increases skills related to prealgebra, embodying computational skills that would relate to problem execution as well as problem representation (Fuchs et al., 2014). This demonstrates the overwhelming benefit of learning mathematic content through the process of problem-solving representation at all stages and not always separating them into different silos.

Problem representation incorporates a combination of a variety of processes, five of which have been found to have significant impacts specifically on word-problem outcomes: number understanding, reasoning, language comprehension, working memory, and attentive behavior (Fuchs et al., 2016). First, number understanding encompasses general understanding of number sense, meaning and effects of different operations, equivalent expressions, and fluency with basic counting strategies (Barrera-Mora & Reyes-Rodriguez, 2019). Controlled for achievement and cognitive measures, high levels of number sense correlate with strong problem-solving skills (Tolar et al., 2016). Furthermore, research shows that a student's computation ability directly impacts how they approach problem-solving strategies (Wang et al., 2016). Young students with poor skills in number sense and numerical computations utilize strategies based on lived experiences compared to older students who choose strategies rooted in arithmetic algorithms, which are viewed as more efficient (Wang et al., 2016).

The second process that impacts problem-solving outcomes is reasoning. Weinstein et al. (2011) teach that a strategic learner is in control of their skill, will, and self-regulation when it comes to academic learning, and they utilize rehearsal, elaboration, and organization learning strategies to improve in those areas. Effective mathematic problem solving involves mastery of and the fluid transition among learning strategies. Strategies that involve deeper or more complex cognition, such as those within elaboration, have been proven to be more effective than those that solely rely on rehearsal strategies for mathematical problem solving (Kikas et al., 2020). Math reasoning also embodies much that has been previously outlined related to self-regulation. For example, well-developed math reasoning skills include tasks such as attempting multiple strategies which is a component of the performance phase of persevering when a tried concept produces a known wrong answer and determining a new strategy to address the math problem at hand.

The third process that impacts problem-solving outcomes is language comprehension. As students engage in the fluid process of creating and testing solutions to a mathematic problem, it is imperative that the student has a strong language base from which to be able to understand, question, and hypothesize issues related to the presented situation. This mental load regarding language use in problem solving relies on more complex cognitive processes than other areas of math achievement; Fuchs et al. (2016) found that language comprehension skills had a significant effect on fourth-graders' math problem-solving but not on prealgebra computations.

The fourth process that impacts problem-solving outcomes is working memory. Research has shown that strong working memory aids in problem solving even when controlled for processing speed and reading ability (Swanson et al., 2008). Students lean on working memory much more heavily in elementary years as more of their time is consumed with computation steps not yet fully automatic (Swanson et al., 2008). As students become more fluent with math facts and computations, working memory is freed to tackle more complicated aspects of mathematic problem solving (Villeneuve et al., 2019). Students who demonstrate weaker working memory skills in early elementary years and are identified as at risk for math problem solving difficulties tend to remain at risk for the same difficulties years later (Swanson et al., 2008).

The fifth and final process that impacts problem-solving outcomes is attentive behavior. Attention focusing is another component of the social cognitive model of SRL previously reviewed (Zimmerman, 2000). This model of SRL involves both an individual's behavioral skill to manage environmental stimulation and their personal motivation to engage those skills in the task at hand (Efklides et al., 2018). Therefore, both external communication skills and metacognition related to the problem-solving process are impacted by an individual's level of attentive behavior.

When looking at results of studies implementing mathematics problem solving instruction and interventions, average achieving students showed stronger problem-solving processing skills than students with learning disabilities or low-achieving students; skills in paraphrasing and visual representation processes were shown to be significant predictors of overall math problem solving performance (Krawec, 2014). Students who were good at problem solving used more strategies and means of representation than those who were not as accurate, meaning that students with learning disabilities often have fewer problem-solving strategies in their toolbox (Montague & Applegate, 2000). Because of this, there is a need to explicitly teach students with learning disabilities how to utilize diagrams of all types prior to introducing cognitive strategy instruction models (van Garderen & Scheuermann, 2015). On top of ensuring understanding of how to use proper diagrams and tables, teaching and modeling strong mathematical vocabulary and dialogue is imperative for stronger outcomes (Hord & Marita, 2014; Powell & Fuchs, 2014; Xin, 2008). The most organic approach to teaching and modeling vocabulary and dialogue involves learning within a context of real-world problems. If students have been prepared appropriately, real-world problem-solving projects are just as effective at reinforcing math problem solving skills with students with disabilities than with those that do not have disabilities (Meyer & Diopoulous, 2002). Unfortunately, special education teachers could benefit from better training in how to effectively teach problem solving to their students with disabilities (van Garderen, 2008).

Self-Regulated Learning and Mathematic Problem Solving

Thankfully, specific methods of implementing SRL within a mathematics-focused learning environment are known to support and increase mathematical reasoning: introducing new materials, self-directed metacognitive questioning, practicing, reviewing additional material, reaching mastery on cognitive processes or strategies involved, verifying accuracy of performance based on corrective feedback, and enrichment or continued practice and use of skill (Mevarech et al., 2018). As can be seen in some of the language utilized to describe components of problem representation, the skills involved in mathematic problem solving naturally lend themselves to aspects of self-regulation. Self-regulation is a component of executive functioning that is correlated with students' mathematic problem-solving struggles (Babakhani, 2011). Specifically, students struggle with determining when to stop trying an unsuccessful strategy, with modifying or generalizing strategies, or with choosing a different strategy (Montague, 2006).

These cognitive processes fall under the SRL concept of metacognition. Simply stated, metacognition defines what a learner knows about and how they manipulate their own thinking and cognitive strategies (Winne, 2018). Research shows that having a growth mindset (versus fixed) and higher levels of self-efficacy and self-confidence relate to increased levels of learning math (Choi & Walters, 2018). Oftentimes a result of having a growth mindset or strong self-efficacy and confidence is increased persistence towards task completion. Due to the more complex nature of math problem-solving tasks, students may need to put forth more effort in order to complete the task, and therefore benefit from higher levels of persistence (Kikas et al., 2020). Even when controlling for intelligence, metacognitive skills predict mathematic problem-solving performance (Berger, 2009). The more awareness and control students have of their cognitive process, the greater the likelihood of their having success in mathematic problem solving (Berger, 2009; Cozza & Oreshkina, 2013; Kazemi et al., 2012).

A hurdle to ideal problem-solving performance that intersects with SRL is emotional regulation as well. Higher levels of worry and math anxiety have been associated with poorer problem-solving performance (Ramirez et al., 2016; Schmitz et al., 2019; Trezise & Reeve, 2018). Ramirez et al. (2016) specifically found that higher math anxiety directly impacts the use of or adequate choosing of advanced problem-solving strategies with elementary-aged students, and Schmitz et al. (2019) found the same correlation with adolescents. Additional correlations have been found among elementary students, poor strategy selection, and high math anxiety specifically in students with higher levels of working memory (Ramirez et al., 2016). Working memory can potentially serve as a partial mediator between problem solving outputs and math anxiety (Ramirez et al., 2016). This could be important to look at with adolescent populations as well. Overall, students' abilities to regulate emotions tied to cognitive processes impacts multiple aspects of the learning process and are another component to consider when observing characteristics of students with potential mathematics disabilities (Tornare et al., 2015).

Although risk factors for students struggling with self-regulated learning and math performance have been identified, even students with cognitive disabilities have been shown to demonstrate metacognitive behavior through math problem solving (Erez & Peled, 2001). The presence of metacognitive activity does not necessarily indicate productive or accurate strategies but does support that students with disabilities can learn through instruction geared toward shaping these specific skills (Rosenzweig et al., 2011). Many studies currently call for increased

SRL measures within math problem solving research, including use of overt or online methods such as traces or think-aloud protocols (Axe et al., 2019; Jitendra & Star, 2011).

The Institute of Education Science has compiled a practice guide report specifically addressing strategies that are evidence-based and focused on improving mathematical problem solving for students in grades 4 through 8 (Woodward et al., 2012). The following recommendations align with research previously reviewed in the realm of mathematics and SRL: assist students in self-monitoring and self-evaluation through the problem-solving process, explicitly teach correct use of visual representations, expose students to a variety of problem-solving strategies, and assist students with understanding and expressing mathematical concepts and notation (Woodward et al., 2012).

Target Population

There has been a range of research pursuing best instructional methods and interventions targeting math problem solving for students in primary and secondary school. It is obvious that the formation of these interventions has been appropriately rooted in the knowledge of critical components of the mathematic problem-solving process and specific characteristics of students who struggle with mathematic problem solving. Although it is encouraging that such a variety of approaches ranging from token economies to computer-based tutoring have been tried, it is not well understood how effective these interventions are for middle school students with disabilities specifically. Similarly, in the SRL research, middle school students are at a critical juncture in early adolescence regarding increasing complexities of academic concepts and increasing needs to cognitively self-regulate throughout their environments (Dent & Koenka, 2016). The presence of a learning disability introduces an additional hurdle for mathematic achievement during middle school as characteristics such as lower reading comprehension (Krawec, 2014), general academic skill deficits across content areas (Mazzocco & Myers, 2003), and weaker executive functioning – particularly working memory (Mazzocco & Myers, 2003; Swanson et al., 2008), inhibit the mathematic problem-solving process.

Purpose of the Study

Students with disabilities are at a disadvantage regarding achieving mastery of mathematic problem solving due to the complexity of their learning profiles (Mazzocco & Myers, 2003). There are many research-based interventions over the past two decades that target mathematic problem solving for students with disabilities (Alter, 2012; Jitendra et al., 2002; Krawec et al., 2013; Ma et al., 2014), many of which do include components of self-regulated learning explicitly (Krawec & Huang, 2017; Montague et al., 2014; Morin et al., 2017; Rosenzweig et al., 2011). This study aims to determine the effect of including an SRL component in mathematic problem-solving interventions for middle school students with disabilities as measured by the results from high-quality methodologically sound single-case design research studies. SRL components will be operationally defined to include terms related to SRL subprocesses involving cognition and metacognition as follows: metacognitive strategies, cognitive strategies, cognitive strategy instruction, schema-based instruction, self-monitoring, self-evaluation, and goal setting (Dent & Koenka, 2016). Data related to effect sizes of interventions for single-case design studies will be analyzed alongside sample size, types of disabilities represented within the sample, delivery location, provider of intervention, specific math content involved, and presence of an SRL component and how it was measured if applicable.

Meta-Analysis

The intention of a meta-analyses is to minimize the limitations present in individual studies by quantifying the effect size of a single independent variable across multiple independent studies (Fraenkel et al., 2014). The ability to apply statistical analysis within the structure of a literature review minimizes potential subjectivity compared to traditional literature reviews (Fraenkel et al., 2014). Synthesizing effectiveness of independent variables through meta-analyses allow practitioners to make more informed decisions (Mikolajewicz & Komarova, 2019). In an application to the field of education, this meta-analysis will provide educators synthesized data to make informed choices for how to best impact student learning in the classroom or create systems-wide decisions related to student services.

The topic for this meta-analysis combines two concepts that are each most prevalent in the separate fields of psychology and education. As it is understood that there may not be a plethora of studies fitting the inclusion criteria, searching more databases across both fields is necessary to ensure all relevant studies are included. After an initial list of articles is confirmed, an ancestral search will be completed to ensure the inclusion of as many articles as possible. Study quality will be addressed using the Council of Exceptional Children's (CEC; 2014) Standards for Evidence-Based Practices in Special Education.

Research Questions

The literature includes reviews looking at effectiveness of interventions involving self-regulated learning theories on overall academic achievement (Dent & Koenka, 2016; Wang & Sperling, 2020) and those looking at effectiveness of interventions involving mathematic problem solving without an emphasis on SRL components (Cook et al., 2020; Peltier et al., 2018; Zheng et al., 2012), but there has not been a meta-analysis of interventions that target mathematic problem-solving skills specifically with self-regulation components for students with disabilities. This study will add to the literature by providing further insight as to the effectiveness of interventions for middle school students with disabilities struggling with mathematic problem solving from the past two decades. The following questions will be investigated by the proposed study:

1. What are the characteristics of effective mathematic problem-solving interventions with self-regulated learning components for middle school students with disabilities?
2. How many CEC quality indicators do each of the included studies meet?
3. What is the effect of mathematic problem-solving interventions with self-regulated learning components on mathematic problem-solving performance for middle school students with disabilities?
4. Do effect sizes vary among problem solving interventions with self-regulated learning components for students with different types of disabilities?

CHAPTER 2

Method

This chapter details the method used to complete a systematic review and meta-analysis of the literature between 2000 and 2021. In this study, twenty-two single case designed studies were reviewed and analyzed for effect sizes related to effectiveness of math problem-solving interventions with middle school students with disabilities. The chapter includes a description of the researcher, independent reviewers, procedures used to conduct the literature review, eligibility inclusion and exclusion criteria, and inter-rater agreement measures. The chapter also includes processes and analysis methods for computing the overall effect of math problem solving interventions that include self-regulation components on students with disabilities' problem-solving performance within middle school grades.

Researcher

The researcher was a doctoral candidate in the field of school psychology with four years' experience teaching as an intervention specialist to middle school students with disabilities. Her teaching license was in mild-moderate educational needs and she met highly qualified teaching status for English language arts and mathematics by the state department of education's licensure standards. She also held a license endorsement for teaching reading and was completing a full-time year-long internship providing school psychological services in an urban school district.

Independent Raters

Three different independent raters were involved in conducting inter-rater agreement (IRA) observations throughout the completion of the study. All three were graduate students in a school psychology program, two of whom were in their third year of study and one of whom was in his second year of study. One had prior experience with coding procedures for research. All three students completed training prior to completing each step of coding to ensure the highest degree of fidelity possible. The first step of the systematic search was split among all three coders. All IRA steps after the initial database search and abstract screening were completed by two independent coders. Additional information about training sessions is included at the end of this chapter.

Inclusion Criteria

Studies that were included in this review and analysis had to be published in English and have been published in a peer-reviewed journal or within a dissertation database between 2000 and 2021. The publication of Zimmerman's (2000) social cognitive approach was used as general metric for limiting studies published prior to this date. The independent variable in the studies had to consist of a mathematic problem-solving intervention. Mathematics problem solving is defined as tasks with math-related word problems where there could be more than one correct pathway to find the correct answer. The studies had to include a dependent measure of mathematic problem-solving performance. Studies needed to include students in middle school within the sample. This was defined as students in 6^{th}-8^{th} grade or 11-14 years old with grade being a priority over age (e.g., a 15-year-old 8^{th} grade student could be included). The study must have reported the presence of a disability with qualification measures noted for at least one student in the sample for the study to be included. Qualification measures could be as simple as stating the student met state criteria to receive special education services and/or detailed information from within the student's Individualized Education Plan (IEP) and evaluation results on intelligence and academic assessments. Only intervention data involving students with a disability were included in the overall effect size analysis.

Finally, interventions within the studies needed to include a component of self-regulated learning (SRL). Components of SRL were defined to also include terms related to subprocesses of SRL and metacognition as follows: metacognitive strategies, cognitive strategies, cognitive strategy instruction, schema-based instruction, self-monitoring, self-evaluation, and goal setting. Cognitive strategies were operationally defined as any part of the intervention package utilizing modeling, corrective feedback, verbal rehearsal, self-questioning, and/or cuing with direct instruction in math problem-solving techniques (e.g., paraphrasing, visualizing, detecting relevant information, locating the question, hypothesizing, estimating, labeling, and/or checking; Krawec & Huang, 2017). Meta-cognitive strategies were operationally defined as tasks that involve thinking about one's own thinking or performance. Schema-based instruction (SBI) is an accepted intervention design that involves cognitive strategies. SBI was operationally defined as instruction in determining problem type and in creating accurate graphic representations to support accuracy during problem solving. Self-monitoring was operationally defined as students checking progress through specific questions or working through steps using a visual checklist. Self-evaluation was operationally defined as a dedicated time at the end of intervention sessions to review and discuss performance on the most recent task compared to previous performance. Goal setting was operationally defined as a dedicated time to look at past performance and set a specific score or goal for future performance. Studies met inclusion criteria if at least one of these SRL components was present within the intervention.

To be included, studies also had to implement a single-case research design (SCRD) and report an effect size (e.g., Tau-U) or the data needed to compute an effect size (i.e., legible graph of results). Systematic reviews, meta-analyses, correlational studies, and qualitative studies were excluded. Studies found were evaluated for quality using the Council of Exceptional Children's (CEC) quality indicators (CEC, 2014). The CEC quality indicators were used over other study quality ratings (e.g., What Works Clearinghouse standards) due to the specific design of the indicators to assess dynamic interventions versus more broad educational programs and curriculums (Cook & Cook, 2011). The CEC quality indicators (2014) looked at eight different categories and each category has a different number of components to evaluate: context and setting (1 component), participants (2 components), intervention agent (2 components), description of practice (2 components), implementation fidelity (3 components), internal validity (9 components), outcome measures/dependent variables (6 components), and data analysis (3 components). Some components within certain indicators only applied to group design studies while others only applied to single-case design studies. The context and setting indicator critiqued descriptions regarding the context and setting of intervention both within a school setting and within a bigger geographic context. The participants indicator critiqued descriptions provided of participant demographic information and participant disability, including how this had been determined. The intervention agent indicator critiqued descriptions provided of the interventionist and any necessary training or qualifications those individuals held or needed. The description of practice indicator critiqued descriptions provided of intervention procedures and materials. The implementation fidelity indicator critiqued study reports of adherence to intervention steps using observable measures, reports of duration and frequency of intervention, and reports of adherence or duration data consistently throughout all phases of the study and across all interventionists. The internal validity indicator critiqued descriptions of how the researcher manipulated the intervention, descriptions of baseline conditions, and lack of exposure to intervention prior to implementation. This indicator also included three components specifically for group design studies: clear description of assignment to groups, measure of overall attrition, and measure of between-group attrition. Another three components critiqued details specifically for single-case design studies: included at least three points of experimental effect, all baseline conditions included at least three data points that established a pattern of steady state responding, and that the design controlled for common threats to internal validity. The outcome measures/dependent variable indicator critiqued importance of outcomes through a social lens, description of measurement for dependent variables, report of results for all mentioned dependent variables, suitability of frequency of outcome measures (at least three data points per phase), and level of interobserver reliability measures. One additional component for this indicator was for group designs only: adequate evidence of validity for content or construct. The final indicator, data analysis, included two group-design specific components and one single-case design specific component. The group-design specific components critiqued statistical techniques reported for group comparisons and the report of an effect size or the information needed to compute one. The single-case design specific component critiqued the presence and general clarity or quality of outcome data in the format of a graph.

All studies that met at least 80% of the components within the quality indicators for single-case design studies were included in the review and analysis. A quality indicator matrix formatted in a Microsoft Excel file designed and made available by Lane et al. (2014) was used to record whether studies meet the criteria for each category. Only the data for participants who met inclusion criteria was considered during the coding process.

Exclusion Criteria

Studies were excluded if they were published in a different language other than English and if they were not published in a peer-reviewed journal or included in a dissertation database. Articles were excluded if they were published or completed prior to 2000. Studies were excluded if the independent variable did not involve a math problem-solving intervention. For example, studies were excluded if the intervention was math-related but focused only on computations instead of involving math problems within the context of a word problem. Studies were excluded if the dependent variable did not measure math problem-solving performance. Studies were also excluded if no middle school students, defined as 6^{th} through 8^{th} grade, were included in the sample. Similarly, if no participants were reported to have a disability, then studies were excluded. If studies implemented a group design, they were excluded. Finally, studies were excluded if interventions did not include components of self-regulated learning or if the study did not meet at least 80% of CEC quality indicators.

Search Terms and Data-Based Sources

The following databases were used to search for research studies that met the inclusion criteria: APAPsychINFO, Academic Search Complete, ERIC (EBSCOHost Version), Education Research Complete, and Psychology & Behavioral Sciences Collection. The inclusion of all of these databases allowed for a thorough search in a wide array of sources spanning the fields of education, psychology, and the social sciences at large. The search terms were as follows: math* AND ("problem solv*" OR "story problem*" OR "word problem*" OR strateg*) AND disab* AND (adolescen* OR "middle school"). Self-regulated learning components of articles were incorporated into the coding procedure, and therefore the search terms allowed for studies that included interventions that may not have mentioned the term self-regulation for middle school samples of students.

Databases were last searched on 03/16/2021. Automatic filters differed by database depending on options available. The APAPsychINFO database used the search terms outlined above and the following filters: limited to 2000-2021, academic journals OR dissertations, English, included school age (6-12 years old) OR adolescence (13-17 years old), quantitative study. The Academic Search Complete and ERIC (EBSCOHost version) databases used the search terms outlined above and the following filters: limited to 2000-2021, academic journals OR dissertations, English. The Education Research Complete database used the search terms outlined above and the following filters: limited to 2000-2021, academic journals OR conference papers, English. The Psychology and Behavioral Sciences Collection database used the search terms outlined above and the following filters: limited to 2000-2021, English.

Search Procedures and Selection Process

All databases were reviewed by the researcher and an independent rater. The researcher reviewed all the databases. The other databases were split across three independent raters. All independent raters were trained on the automation filters and search terms for each database prior to conducting the search for studies. Independent raters conducted the search on assigned databases and screened the references and abstracts of all results for inclusion/exclusion criteria. If details were shared that would exclude the study (i.e., systematic review, high school students, reading intervention, group design, etc.), then the study was excluded. If no exclusionary information was included, but the abstract was still vague, all independent raters were trained to mark the study for inclusion to read the studies in full to determine if they met the inclusion criteria. The researcher calculated the IRA for each database and created a master list, omitting duplicate studies. IRA was then calculated for the master list, both before and after omitting duplicate studies. The next step of the systematic search was for each study to be read in full for the same inclusion/exclusion criteria independently by the researcher and an independent rater. The researcher read all the articles in full. The sample was then split evenly across three independent raters for coding. The lead researcher compiled the IRA calculations and created a final list of included articles.

An ancestral search was completed on all articles that met inclusion criteria. The researcher reviewed the references within all the included articles and identified those that met the inclusion/exclusion criteria for this review (e.g., limited to 2000-2021, no reviews or meta-analyses, title involving keys words of math problem solving, intervention, self-regulated learning). The compiled list of references was then cross checked to the original database list of titles reviewed. All duplicate references already identified were removed. Of the original 20 articles, five were randomly chosen for an independent rater to review the reference citations by independently following the same steps as above. Next, the lead researcher identified and read the abstracts from the titles that appeared to meet the inclusion criteria from the ancestral search. Five of the abstracts that appeared to meet the inclusion criteria were randomly selected and given to an independent rater to review following the same steps as the lead researcher. Last, the lead researcher located the articles that met the inclusion criteria based on the abstract review. Both the lead researcher and an independent rater read each article in full and made an independent determination regarding final inclusion for this review. A final list of included articles was created by combining the results of the initial systematic search and the ancestral search.

Data Collection Process

Article PDF files were downloaded and stored in a computer software program called Mendeley on the researcher's laptop and on a Google Drive for the independent raters to access. Separate pre-formatted Excel files housed the systematic search data, CEC quality indicator data, and specific content data. Table 1 outlines the information coded from within each included article. Articles were coded according to the following: alignment with CEC quality indicators; number of participants that met inclusion criteria; grade of participants included; diagnosis/disability qualification; test used to determine diagnosis; race/ethnicity;

Table 1

Data Collected Within Studies During Coding Process

Data gathered by focus area
Quality of study
Alignment with the Council for Exceptional Children's (2014) eight quality indicators
Participants
Number of participants that met inclusion criteria
Grade of participants included
Diagnosis/disability qualification
Test used to determine diagnosis
Race or ethnicity reported
Study details
Brief description of the intervention program
Problem types and/or operations included within context of problem-solving intervention
Dependent variable(s)
Implementation grouping (i.e., one-on-one, small group, or large group)
Interventionist(s)
Setting of intervention within building
Self-regulated learning
Components of self-regulated learning included in the independent variable
Data related to SRL component was considered offline or online
Effect size
Graph of results for dependent variable

implementation grouping (i.e., one-on-one, small group, or large group); brief description of the intervention program; problem types and/or operations included within context of problem solving; dependent variable(s); interventionist; setting of intervention; components of self-regulated learning included in the independent variable and if those components were measured in an offline or online manner; and effect size or data necessary to compute it.

Coding Procedures for Specific Items

For the CEC quality indicators, all articles in the final sample were coded independently by the researcher and an independent rater. The researcher coded all articles, and the sample of articles were split in half for secondary coding to be completed by two independent raters. Each reviewer was trained on the CEC's quality indicators and used the pre-formatted Excel spreadsheet matrix for coding purposes (e.g., Lane et al., 2014). The matrix detailed each subcategory for the quality indicators and prompted reviewers to code either a 0 (unmet) or 1 (met) for each item per study. The researcher and independent raters also cited page numbers and quotes from the article showing evidence where studies met individual indicators and stated a reason for marking that it was unmet if applicable. Once the coding was completed by the researcher and independent raters, the researcher compiled the data into one Excel matrix. The matrix then computed overall IRA by quality indicator, by article, and the number of quality indicators met by study both using an absolute method and an 80% weighted method (Royer et al., 2017). When using absolute coding, studies earned a "1" for each quality indicator in which they met all components within it. If there was a single component that was not met, the study earned a "0" for that indicator when implementing absolute coding. Weighted coding allowed for minor lapses in study design or reporting by providing weighted scores for each indicator (i.e., if 3 of 5 components were met, that indicator would be noted with a weighted score of 60%). When determining studies that could be considered as methodologically sound with absolute coding, studies had to meet all components of all 8 quality indicators. When determining studies that could be considered as methodologically sound with weighted coding, studies had to meet a threshold of at least 80% of components met (overall rating of 6.40 quality indicators met or higher).

The participant and study details of each intervention were coded directly from the studies by the researcher. If only a portion of participants met the grade/age inclusion criteria, only that data was included in the coding and analysis. Grade was weighted higher than age for inclusion; for example, if a 15-year-old (outside the range of inclusion for age) participant was marked as being in the 8th grade (inside the range of inclusion for grade), then that student's data was included in the analysis. The same rules applied for any younger students within the middle-school grade band. For coding of the test used to determine diagnosis, some studies did not list specific assessments. In these instances, notes were made regarding the evidence cited that allowed for inclusion of the participant as an individual with a disability (e.g., receiving special education services, having an IEP, etc.). All remaining coding categories were either present within the article or not. If unaddressed, the coding was left blank and accounted for and reported within the analysis as missing data. Any compatible information for these categories was noted and coded as possible.

The self-regulated learning components were defined as implementation or utilization of any of the following aspects of self-regulated learning: cognitive strategy, goal setting, planning, self-monitoring, self-control, and self-evaluation (Dent & Koenka, 2016). These aspects of self-regulated learning have common threads across a myriad of SRL models and frameworks (Dent & Koenka, 2016). The researcher coded two categories related to SRL, first if any of these areas as detailed by Dent and Koenka (2016) were present, and second, which data was present within the study, either an offline measure (i.e., self-report or structured interview) or an online measure (i.e., traces such as think-alouds or underlining important information that provides a visual snapshot of the internal SRL component; Dent & Koenka, 2016). At least 30% of articles were coded for these components by an independent rater to confirm results.

In order to conduct the meta-analysis, the web-based program Web Plot Digitizer was utilized to extract all individual data points from visual graphs included within the original studies for participants that met inclusion criteria. For studies that included data points on multiple skills within a single graph, the data was pulled apart into two different data sets if there were enough data points within each phase. Multiple intervention phases were all treated as a single intervention subphase for the purpose of data analysis within a multiple-probe design. Additionally, generalization and maintenance probes were not included in the data sets analyzed as not all studies included them or were consistent in how and when these measures were implemented. Data was organized in an Excel sheet with a tab for each study. Each tab included a column for each AB comparison, usually corresponding with a column per participant, that detailed data points and condition assignment (e.g., baseline and intervention). The data sets were then imported into a virtual calculator at http://www.singlecaseresearch.org/ created by Vannest et al. (2016) to calculate the effect size using Tau-*U*. Tau-*U* allows for a quantitative analysis of effect size within single-case research data by combining measures of nonoverlapping data, intervention phase trends, and corrections for unstable baseline trends across multiple studies (Lee & Cherney, 2018). Each participant's baseline was first checked for any significant trends. If these were present, the Tau-*U* calculator was utilized to correct baseline trends prior to contrasting that participant's baseline and intervention phase data. In this study, any baseline trend with a Tau-*U* measure greater than ± 0.2 was corrected as recommended by Vannest and Ninci (2015). Contrasts of baseline to intervention phases were calculated for each participant to obtain an effect size for each participant. The contrast data for all contrasts involved in the study were then combined into a single weighted overall Tau-*U* for the study. Data outputs from the online calculator were saved onto additional tabs within the effect size excel spreadsheet.

An aggregate mean, or overall effect size, of the interventions on math problem-solving performance was calculated using the Jamovi statistical program's meta-analysis features. Using continuous outcomes with pre-calculated effect size data and a random-effects model, the overall effect size output included a forest plot with overall effect size, confidence intervals, and heterogeneity assumption tests. A sensitivity analysis was completed that reran the overall effect size analysis but included only studies that met all eight of the CEC's quality indicators when using absolute coding. With similar results, this analysis increased validity of the reported overall effect size when including studies that met methodological strength using the weighted coding versus the stricter absolute coding.

Additional analyses were completed to determine any potential moderator effects using the IBM® SPSS® Statistics meta-analysis feature by computing subgroup analyses with the Hunter-Schmidt estimator method. The Hunter-Schmidt estimator method is recommended due to demonstrating less bias in analysis data available (Field & Gillet, 2010). Tests of heterogeneity were computed for each potential moderator analysis as well to strengthen support for conclusions drawn from the results. Moderator analyses were conducted for disability category, interventionist, group setting, intervention type, and problem type. The analyses were all conducted on a data set created with each participant's effect size, variance, disability category reported, interventionist, group setting, intervention type, and problem type.

Risk of bias measures were conducted to increase transparency and reliability of the data reported and conclusions drawn from the results. Although digital tools such as the Risk of Bias 2 (Sterne et al., 2019) that are based on the Cochrane Collaboration's guidelines for risk of bias assessment are freely accessible, none of these tools are well-designed for single-case design studies as they focus on randomized-controlled trials. Reichow et al. (2018) proposed a framework based on the Cochrane risk of bias tool that allows for the incorporation of single-case design studies in the risk of bias assessment. This analysis implemented Reichow et al.'s (2018) framework and description of their Single Case Design Risk of Bias (SCD RoB) tool to conduct a risk of bias assessment. A protocol or tool for the SCD RoB was not made available by the researchers, but adequate descriptions of each area of bias and domain were included in the original article (see Appendix A) along with example outputs. The SCD RoB focused on selection, performance, and detection biases. Selection bias included sequence generation and participant selection. Performance bias included blinding of participants and personnel and procedural fidelity. Detection bias included blinding of outcome assessment, selective outcome reporting, dependent variable reliability, and data sampling.

Articles were coded for a low, unclear, or high risk of based in each of the eight domains outlined above based on guidance provided by Reichow et al.'s (2018) SCD RoB framework. The results of each study were then combined to create an overall measure of low, unclear, and high risk of bias per bias domain. These results were then analyzed for potential implications or limitations on the conclusions drawn from the overall results to support a final determination on the certainty of evidence presented in this meta-analysis.

Training for Inter-Rater Agreement and Treatment Integrity

The independent raters that assisted with coding articles for the initial systematic search, with coding articles on CEC's quality indicator checklists, and with coding select content pieces within studies completed multiple training sessions throughout the process of conducting this analysis. At any point any discrepancies were discussed between the researcher and the independent rater until there was 100% agreement. Six different training meetings were held that addressed treatment integrity for the systematic search, coding articles for alignment to the CEC's quality indicators, and coding content items.

The first two training sessions focused on the systematic search. The independent raters and lead researcher reviewed the first 10 articles from the APAPsychInfo database together as a group. The researcher led the training on using search terms and setting appropriate settings and limitations (i.e., viewing the abstract, limiting included years, and including only journals and dissertations/theses). The researcher modeled the process of reading the abstracts for the following criteria: (1) single-case design empirical study, (2) involved middle school students (aged 10-14 or grades 6-8), (3) with disabilities (any disability category as long as the data for students with disabilities is separate from general population data), (4) some sort of intervention related to math problem-solving performance, and (5) a dependent measure related to mathematic problem-solving performance. The next articles were completed as a group with the independent raters leading discussion around whether the article would meet inclusion criteria or not. Finally, the last two articles were reviewed by each of the independent raters independently before coming back together to compare results. There was 100% agreement in this session, so no more articles were reviewed from this database for training purposes.

Prior to coding the sample for alignment with the CEC's quality indicators, two independent raters met with the researcher three times to train and calibrate how to complete the coding accurately. The Excel overview and walk-through guide (Royer et al., 2017) and the paired Check for Understanding (CFU; Common et al., 2017) documents were provided to the independent raters prior to the meeting, along with the four primary studies that had been written as guidance for using the quality indicators and are referenced in the Excel document and training guide. After the initial meeting, each independent rater read the walk-through guide in depth and studied the Excel document independently before taking the CFU. Each independent rater reached a minimum of a 90% pass rate on the CFU (r = 95%-100%; n = 3). In the next meeting, the independent raters practiced independently coding three articles not included in the current study for treatment integrity training on the quality indicators. Across the three articles, the lead researcher had an overall rate of 84.86% agreement with the first independent rater and 88.38% agreement with the second independent rater. After the results were compiled, another training meeting was held to talk through any discrepancies and how to make decisions related to those issues when coding articles for this study.

A final training session was held to calibrate two independent raters on coding content in articles for the presence of a SRL component and if there were data collected in an online or offline manner related to the SRL component. The researcher modeled coding for one of the included articles. Another article was completed together as a group, talking through the decision process with the independent raters. A final article was coded independently and results compared during the session with 100% agreement among raters.

CHAPTER 3

Results

The results of the systematic study selection and meta-analysis are reviewed in this chapter organized by research question. Additional information is included regarding risk of bias and additional potential moderators in addition to type of disability.

Study Selection

Each independent reviewer was provided one or two of the following databases to search for articles that met the inclusion criteria for this systematic review: Psychology and Behavioral Sciences Collection, Education Research Complete, ERIC, Academic Search Complete, and APA Psych Info. Based on an initial search, the databases were split and assigned to reviewers in such a way to create a generally even distribution of titles and abstracts to review. Databases were grouped and then randomly assigned to a reviewer. Each reviewer screened the articles for inclusion by reading the titles and abstracts. The reviewers recorded the search terms in each database separately, and all resulting articles' abstracts were screened for potential inclusion by the researcher and an independent reviewer. The resulting inter-rater agreement (IRA) for screening abstracts was as follows: Psychology and Behavioral Sciences Collection (initial IRA = 86.79%, n = 53); Education Research Complete (initial IRA = 89.38%, n = 113); ERIC database (initial IRA 86.34%, n = 205); Academic Search Complete (initial IRA = 91.13%, n = 124); APA Psych Info (initial IRA = 93.48%, n = 46). The overall IRA was 87.99% (n = 541). The researcher reviewed all the databases and created a master list of the articles that met the inclusion criteria based on the how the abstracts were coded. After removing duplicate articles from subsequent databases, the overall IRA for all screened articles was 91.91% (n = 309). Any discrepancies were resolved with conversations between the researcher and the independent reviewers. Following this procedure, 52 articles were determined to meet initial inclusion criteria and were then read in full.

The 52 articles were divided evenly and read in full across the independent reviewers. The researcher read all the articles and compiled the reviewers' coding decisions into the master list. The IRA agreement for this step was 90.00% and resulted in 20 articles that met inclusion criteria. The next step involved an ancestral screening search of article titles included in the reference sections of the 20 articles that met the inclusion criteria. There were 1,175 total ancestral titles included across the sample of 20 articles. Five articles were chosen at random for an independent reviewer to screen ancestral title citations. Within those five articles, a total of 194 ancestral reference titles (16.4%) were screened by both the independent reviewer and the researcher. IRR for ancestral citation screening was 93.75%. Eighty-eight total article titles were identified as being studies that may meet the inclusion criteria. Of those titles, 13 were identified by only one reviewer, 37 had been previously screened in the original eSearch, and 18 articles were eliminated as duplicate findings to those in the final search. This left 20 new and unique articles to be screened by looking up and reading the abstracts. Abstracts for two out of the 20 article titles could not be located. The lead researcher located the abstracts and PDFs of the remaining 18 article titles and screened each abstract to determine if they should be read in full for final inclusion. An independent reviewer screened five randomly selected studies of the 18 articles in this step (27.8%). The IRA agreement on which articles should move on to be screened by reading in full was 100.0%. Three articles were identified to be read in full, and two met the final inclusion criteria. An independent reviewer read all three articles in full, and IRA was 100.0%. After the ancestral search was completed and the additional 2 articles were added, it was determined that 22 single-case design studies met the inclusion criteria for this meta-analytic review. Table 2 provides a summary of IRA results and Figure 2 depicts a diagram of the search procedures.

Several articles were excluded due to variety of situations. For instance, Zhang, et al.'s (2016) article, Strategic development for middle school students struggling with fractions, was excluded due to the independent variable consisting of solving numerical fraction computation problems only (e.g., 5/8 – 1/3) and not problems written within a story problem context even though real-world problem-solving was a dependent variable. Bundock et al.'s (2021) article, Teaching rate of change and problem solving to high school students with high incidence disabilities at tier 3, was also excluded due to all participants being in high school and outside of the 6^{th}-8^{th} grade range. Cox and Root's (2020) article, Modified schema-based instruction to develop flexible mathematics problem-solving strategies for students with Autism Spectrum Disorder, was excluded because the intervention was implemented at a student's home and at a local library, instead of within the context of a school setting. Finally, Brawand's (2013) dissertation, Proportional reasoning word problem performance for middle school students with

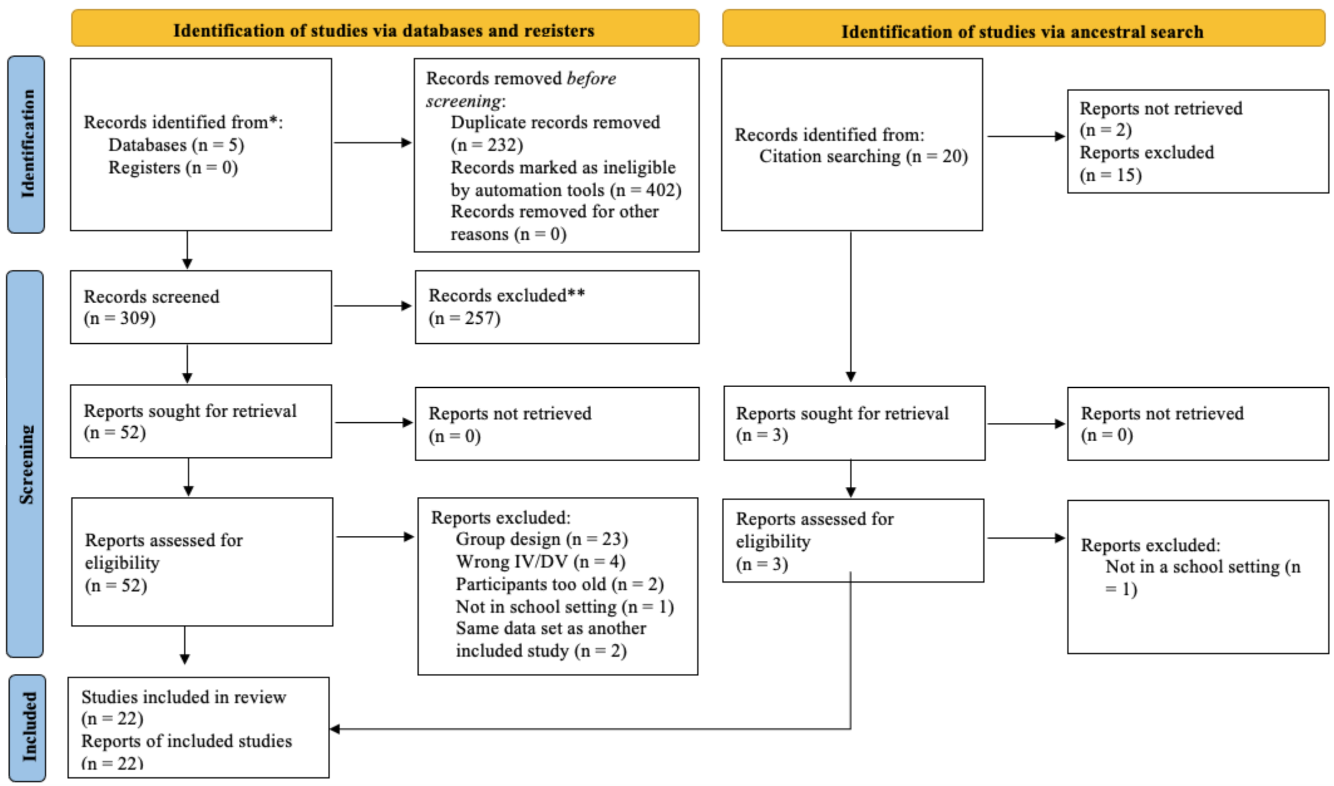
Table 2

Interrater Agreement (IRA) Throughout Systematic Review

Step	Total rated articles	IRA
Initial Database Search, No Duplicates	309	91.91%
Articles Read in Full	52	90.00%
Ancestral Citation Screening	192 (16.4%)	93.75%
Ancestral Abstracts	5 (27.8%)	100.00%
Ancestral Read in Full	3	100.00%

Figure 2

Flow Diagram for Systematic Review



Note. Adapted from "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," by M. J. Page et al., 2021, *The British Medical Journal, 372*(71), p. 1–9 (https://doi.org/10.1136/bmj.n71). Copyright 2021 by PRISMA.

high-incidence disabilities, was excluded because the same data set was used in a published article that also met inclusion criteria (Brawand et al., 2020).

Question 1. Study Characteristics

Table 3 presents a summary of the following study characteristics: participants, setting, interventionist, intervention, and type of problem solving utilized. Across the 22 single-case design studies that met the inclusion criteria, there was a total of 94 students with disabilities in middle school settings. Of the 22 studies, the researcher was the interventionist in sixteen studies (72.7%), a special educator was the interventionist in four studies (18.2%), peer tutors were interventionist in one study (4.5%), and a computer program was used in one study (4.5%). Sixteen studies (72.7%) implemented the intervention package in a one-on-one setting, whereas the other six studies (27.3%) implemented the intervention in a small group setting. Of the 94 participants, 45 participants (47.9%) were identified with learning disabilities, 33 participants (35.1%) were identified with intellectual disabilities, 13 participants (13.8%) were identified with autism spectrum disorder, 1 individual (1.1%) was identified with attention deficit hyperactivity disorder, 1 individual (1.1%) was identified under the educational disability category of Other Health Impairment, and 1 individual (1.1%) was identified with an auditory processing disorder. All participants received special education services according to their state education criteria. Participants ranged in age from 11 to 15 with a mean of 13.2 years of age. Sixty-five participants (69.1%) were male and 29 (30.9%) were female. With regard to race and ethnicity of the 94 participants, 55 (58.5%) were White, 20 (21.3%) were Black, 9 (9.6%) were Hispanic, 2 (2.1%) were mixed race, and 1 (1.1%) was Middle Eastern. Two studies involving 7 participants (7.4%) did not list race/ethnicity demographic information. The analysis included 112 different

Table 3

Study Characteristics

Study	Participants		Setting		Interventionist	Intervention	Type of problem solving
	n	Disability*	Geographic*	Building			
Bouck & Long (2020)	3	ID, AU, ID	Midwestern USA, Rural public school	Private separate space	1-on-1 Researcher	Use of a schematic diagram and a system of least prompts with four levels	Percent change: Ratios/fractions/cost
Bouck et al. (2021)	4	ID, AU, LD, ID	Rural public school	Hallway outside classroom	1-on-1 Researcher	Use of a schematic diagram, taught via explicit instruction with fading, and a system of least prompts with three levels	Percent change: Finding total cost of meal with tip
Bouck et al. (2017)	4	ID, LD, ID, LD	Midwestern USA, Rural public school	Hallway outside classroom	1-on-1 Researcher	CRA instructional sequence for solving change-making with coins problems. In the concrete phase, students used plastic coins and paper bills to solve the problems. In the representational phase, students drew images (i.e., circles with numbers [5, 25] or rectangles) to represent coins and bills to solve the problems. Finally, during the abstract phase, students solved the problems without any support"	Group, Change, Compare with Addition/Subtraction: Change-making with coins

Table 3 (continued)

Study	Participants		Setting		Interventionist	Intervention	Type of problem solving
	n	Disability*	Geographic*	Building			
Brawand et al. (2020)	9	AU, LD, APD, ADHD, LD, LD, LD, LD, LD	Northeastern USA, Suburban public school	Private separate space	Small group Researcher	Schema-based instruction (SBI) for proportional reasoning word problems within small groups	Vary and multiplicative comparison: Proportions
Browder et al. (2012)	4	All moderate/ severe ID	Southeastern USA, Urban public school	Special education classroom	Small group Special Education Teacher	Combination of read-aloud of a math word problem, a graphic organizer, and task analytic instruction in the steps to solve the problem	Not reported; Curriculum-based
Browder et al. (2018)	6	All ID	Southeastern USA, Urban public school	Special education classroom	Small group Special Education Teacher	Modified schema-based instruction (MSBI) including explicit instruction with systematic feedback/error correction, metacognitive strategy instruction, task analysis, graphic organizers, self-monitoring, and manipulatives	Group, change, compare with addition/subtraction
Freeman-Green et al. (2015)	6	All LD	Southeastern USA, Private school for LD or ADHD	Unclear	Small group Researcher	SOLVE strategy in the area of algebra, involving explicit instruction, promotion of metacognition in the context of problem solving, and uses consistent process as other research-based mnemonic strategies for math problem solving	Curriculum-based; Combination of problem types involving addition/ subtraction and multiplication/ division

Table 3 (continued)

Study	Participants		Setting		Interventionist	Intervention	Type of problem solving
	n	Disability*	Geographic*	Building			
Jitendra et al. (2002)	4	All LD	Northeastern USA, Suburban public school	Special education classroom	1-on-1 Special Education Teacher	Schema-based instruction (SBI) is a representational strategy that focuses on identifying problem patterns or structures through the use of schemata diagrams to map information/highlight sematic relations in the problem to increase accuracy in solving	Vary and multiplicative comparison
Ley Davis (2016)	4	All moderate/ severe ID	Southeastern USA, Urban public school	Private separate space	1-on-1 Peer tutors without disabilities	Peer-mediated schema-based instruction (SBI) for students with moderate/severe ID during peer-tutor delivered instruction	Group, change, compare with addition/subtraction
Maccini & Ruhl (2000)	3	All LD	Central Pennsylvania, Public school	Private separate space	1-on1 Researcher	STAR, an algebra problem-solving strategy (Maccini, 1998), includes (a) a concrete, semi-concrete, abstract (CSA) instructional sequence; (b) general problem-solving strategies; and (c) self-monitoring strategies	Group, change, compare with addition/ subtraction
Na (2009)	4	LD, LD, OHI, LD	Not listed	Tutoring area	1-on-1 Researcher	Schema-based instruction (SBI)	Vary and multiplicative comparison
Root, Cox, Davis, & Hammons (2020)	3	AU, AU, ID	Southeastern USA, Suburban public school	Private separate space	1-on-1 Researcher	Modified schema-based instruction (MSBI) paired with an iPad calculator, video anchors of specific skill being used in community locations,	Percent change: Finding total cost of item with coupon

Table 3 (continued)

Study	Participants		Setting		Interventionist	Intervention	Type of problem solving
	n	Disability*	Geographic*	Building			
						and goal setting with self-graphing of progress. Follow-up prompted participants to determine if a given amount of money was enough to make the purchase	
Root, Cox, Gilley, & Wade (2020)	3	ID, AU, AU	Southeastern USA, Public school	Private separate space	1-on-1 Researcher	Virtual-Representational-Abstract-Integrated (VRA-I) framework presenting multiple representations (e.g., concrete/virtual with representation or representation with abstract)	Vary and multiplicative comparison
Root, Cox, Saunders, & Gilley (2020)	3	ID, ID, AU	Southeastern USA	Private separate space	1-on-1 Researcher	Modified schema-based instruction (MSBI) components further individualized using universal design for learning (UDL) framework to minimize barriers for problem solving tasks	Percent change: Increase and decrease
Root et al. (2018)	2	Both AU	Southeastern USA, Private school for AU	Private separate space	1-on-1 Researcher	Modified schema-based instruction (MSBI) addressing algebraic reasoning and problem solving with quantities above ten and discrimination of problem types	Group, change, compare with addition/ subtraction

Table 3 (continued)

Study	Participants		Setting		Interventionist	Intervention	Type of problem solving
	n	Disability*	Geographic*	Building			
Root et al. (2017)	3	All moderate ID	Southeastern USA, Urban public school	Private separate space	1-on-1 Researcher	Modified schema-based instruction (MSBI) with a calculator	Group, change, compare with addition/ subtraction
Saunders et al. (2018)	3	All moderate ID	Southeastern USA, Urban public school	Private separate space	1-on-1 Researcher	Use of video prompting to teach and reinforce strategies (finger counting) for word problems involving real-world situations simulated in a video clip	Group, change, compare with addition/ subtraction
Schaefer Whitby (2013)	3	All AU	Central Florida, Public school	Private separate space	1-on-1 Researcher	Solve It! Problem Solving Routine curriculum (Montague, 2003). The curriculum consists of teaching students seven cognitive strategies (read, paraphrase, visualize, hypothesize, estimate, compute, and check) and three meta-cognitive strategies (self-management, self-questioning, and self-evaluation)	Not reported; Curriculum-based
Scheuermann et al. (2009)	14	All LD	Mideastern US, Charter school for LD	Private separate space	Small group Researcher	Explicit Inquiry Routine (EIR) involves three instructional steps: explicit sequencing (pre-determine chunking of main concept and teach in small individual instructional sessions), scaffolded inquiry (answer three set guiding questions with instructor within	Combination of problem types involving addition/ subtraction and multiplication/ division

Table 3 (continued)

Study	Participants		Setting		Interventionist	Intervention	Type of problem solving
	n	Disability*	Geographic*	Building			
						each step), and systematic use of various modes of illustration (concrete, representational, and abstract)	
Sheriff & Boon (2014)	3	All ID	Southeastern USA, Rural public school	Special education classroom	Small group Special Education Teacher	Computer-based graphic organizers to assist in solving one-step word problems	Combination of problem types involving addition/ subtraction and multiplication/ division
Shin & Bryant (2017)	3	All LD	South central USA, Private school for LD	Private separate space	1-on-1 Computer program	Multi-component computer-assisted instructional program (CAI), Fun Fraction. The multi-components include cognitive and metacognitive strategies, virtual manipulatives, and explicit, systematic instruction	Combination of problem types involving addition/ subtraction and multiplication/ division
van Garderen (2007)	3	All LD	Mid-Hudson region of New York state, Public school	Private separate space	1-on-1 Researcher	Strategic instruction in diagram generation and use with three instructional phases: diagram instruction, one-step word problem strategy instruction, and two-step word problem strategy instruction	Combination of problem types involving addition/ subtraction and multiplication/ division

Note. ADHD = Attention deficit hyperactivity disorder; APD = Auditory processing disorder; AU = Autism; DS = Down's syndrome; ID = Intellectual disability; LD = Learning disability; OHI = Other Health Impairment.

AB phase (A: baseline; B: intervention) contrasts. All contrasts included a total of 1,556 data points extracted from the graphs provided within each published study.

Interventions

The independent variables implemented throughout the studies included in this analysis involved a variety of approaches. The researcher sorted the interventions into six categories based on similarities: variations on general cognitive/metacognitive strategy routines, specific concrete-representation-abstract (CRA) instructional routine variations, schema-based instruction, modified schema-based instruction, instruction specifically involving schematic diagrams, and technology-based intervention approaches.

General Cognitive/Metacognitive Strategy Variations. Three studies (13.6%) involved variations on general problem-solving routines. For instance, Browder et al.'s (2012) study focused on effects of reading problems out loud paired with a visual that contained picture symbols and graphic organizers that could be modified using Velcro numbers for students who were not able to write with dry erase markers. It also included a specific task checklist of 5-7 steps that students were taught and used independently. This study included data from four different math units (e.g., geometry, algebra, data, and measurement). Only the data for the algebra unit was used in this meta-analysis. In Schaefer Whitby's (2013) study, the curriculum Solve It was explicitly taught to students with autism. It incorporated seven cognitive strategies (read, paraphrase, visualize, hypothesize, estimate, compute, and check) and three meta-cognitive strategies (self-management, self-questioning, and self-evaluation). In Freeman-Green et al.'s (2015) study, students were explicitly taught to use a mnemonic (SOLVE) as an effective strategy to complete math problem solving through modeling, verbal practice, controlled practice and feedback, and advance practice and feedback.

Concrete-Representation-Abstract Sequence Variations. Four studies (18.2%) involved teaching math problem-solving using variations of concrete-representation-abstract (CRA) problem-solving interventions. For instance, in Maccini and Ruhl's (2000) study, students were taught a mnemonic (STAR: Search the word problem, Translate the words into an equation in picture form, Answer the problem, Review the solution). Each phase in the mnemonic involved a three-step task analysis. Students completed the first two steps multiple times through representing the problem first with physical manipulatives and then drawing a picture of the representation. Finally, students completed all four steps of the mnemonic by writing and solving an algebraic equation. Instructional strategies such as modeling, advance organizers, guided practice, and independent practice were implemented throughout. In Scheuermann et al.'s (2009) study, a variation of a problem-solving routine called Explicit Inquiry Routine (EIR) was implemented with students. This intervention involved chunking the main mathematical concept – in this case, one-step word problems – into smaller instructional lessons prior to beginning instruction, and guiding students through solving the problem by asking sets of questions using a think-pair-share model. Throughout each step students were encouraged to represent the problem in multiple ways using manipulatives, pictures, and algorithms. In Buock et al.'s (2017) study, a concrete-representational-abstract approach was used to teach students to make change with coins by first using fake money, then drawing pictures to represent the problem. In the last step, students solved the problem using algorithms or mental math. Root, Cox, Gilley, and Wade's (2020) study implemented a virtual-representation-abstract (VRA) sequence with students who have significant support needs. This sequence involved all aspects of making different types of models that the other studies did, but the concrete step used technology and other resources to allow students to create a representation of the problem by virtually manipulating objects during the first phase.

Schema-Based Instruction (SBI). Four studies (18.2%) implemented schema-based instruction. Jitendra et al.'s (2002) study used a visual representation strategy focusing on recognizing problem patterns. This involved using set diagrams to highlight important information on multiplicative and vary problem types (e.g. Renae earned $15 for each day she babysat. She worked for 5 days. How much money did she earn?). Similarly, Na (2009) also implemented schema-based instruction with one-step multiplication and division word problems. Ley Davis (2016), on the other hand, taught peers without disabilities to deliver schema-based instruction in a one-on-one setting to peers with moderate-to-severe intellectual disabilities. The type of problem involved change problems with addition and subtraction. Most recently, Brawand et al.'s (2020) study considered the effects of schema-based instruction on proportional reasoning word problems for students with learning disabilities, other health impairment, and Autism when implemented within small groups.

Modified Schema-Based Instruction (MSBI). Five studies (22.7%) implemented a modified schema-based instruction. MSBI incorporates all essential components of SBI while adding evidence-based components that support students with moderate intellectual disabilities and/or significant communication and reading needs. For instance, Root et al.'s (2017) study provided students with a calculator in addition to the MSBI components of instruction, task analysis and read aloud to teach students to solve problems with money quantities larger than 10. In Browder et al.'s (2018) study, MSBI was implemented with some problems being presented through video-simulation to determine how well students solved math word problems when required to discriminate among problem types independently. Set graphic organizers for each problem type were provided and students could choose which organizer was needed to solve each problem. Root et al.'s (2018) study used MSBI to teach students how to solve problems containing quantities larger than 10 with word problems involving addition and subtraction. In Root, Cox, Saunders and Gilley's (2020) study, MSBI was implemented along with strategically chosen components of the universal design for learning (UDL) framework to minimize barriers throughout the problem-solving process for students with intellectual disabilities and autism. Root, Cox, Davis, and Hammond's (2020) study implemented a version of MSBI by conducting a replication of Root et al.'s (2018) study. The replication tweaked the measurement strategy in order to be more sensitive to student's understanding of payment (e.g., instead of saying if there was enough money, students had to provide a number for the next whole dollar amount needed to cover the full cost).

Diagram-Based Approaches. Three studies (13.6%) instructed students specifically on creating and using strategic diagrams during the problem-solving routine. For instance, in van Garderen's (2007) study, students with learning disabilities were explicitly taught how to create diagrams that were effective in supporting an accurate solution to a given one- or two-step word problem. This involved explicit strategy instruction for each type of word problem once students learned appropriate techniques for generating useful diagrams. Buock and Long (2020) provided a specific diagram and taught students how to complete it using a task analysis checklist and a strategic four-step system of least-to-most prompts including gesture, indirect verbal, direct verbal, and modeling while solving percent change problems (cost of item after a discount). Building upon that research, Buock et al.'s (2021) provided a specific diagram, task analysis checklist and a strategic three-step system of least-to-most prompts including gesture, indirect verbal, and modeling while solving percent change problems (e.g., finding total cost of meal with tip).

Technology-Based Approaches. Finally, three studies (13.6%) utilized technology-based interventions to increase problem-solving skills for students with disabilities. Saunders et al.'s (2009) study used video clips of simulated real-world situations and video prompting of math problem-solving steps to teach students with moderate intellectual disabilities to solve addition and subtraction change problems. In Sheriff and Boon's (2014) study, special education teachers were trained to work with students in small groups on one-step addition, subtraction, or multiplication problems by providing graphic organizers virtually versus paper formats. In Shin and Bryant's (2017) study, a computer program called Fun Fractions was implemented to integrate cognitive strategies, feedback, and virtual manipulatives to teach math problem solving to students without additional instruction from teachers or adults.

Problem Types

Real-world mathematic word problems involve an underlying structure that can be recreated in different contexts. Oftentimes problem-solving instruction revolves around specific structures of word problems to minimize confusion among students. There were a variety of problem types that were taught in the included studies. These were grouped into percent change problems, group/change/compare problems involving addition and subtraction problems, vary and multiplicative comparison problems, unlisted curriculum-based problems, and combinations of group/change/compare and vary multiplicative comparison problems. Seven studies (31.8%) involved group, change, and compare addition and/or subtraction type problems. Four studies (18.2%) consisted of percent change problems involving ratios, fractions, and cost, and four more studies (18.2%) focused solely on vary or multiplicative comparison type problems. Five studies (22.7%) contained a combination of problem types involving all four operations. Three studies (13.6%) reported problems as being curriculum-based.

Primary Dependent Variable

The primary dependent variable for all studies was problem solving accuracy as defined by completing steps of an intervention or strategy and arriving at a correct answer for given word problems independently. The number of steps included in the outcome measure was not consistent across all studies, ranging from 1 step (solely the correct answer) to 12 steps in the form of a task analysis checklist. The most frequent measurement was 1 step (9 studies; 40.9%). Both 5-step and 12-step task analyses – including final correct answer – were the next most common levels of measurement represented in 3 studies each (13.6%).

Secondary Dependent Variables, Social Validity, and Generalization Measures

Secondary Dependent Variables. Table 4 provides a summary of the secondary dependent, social validity, generalization, and maintenance measures that were incorporated across the sample. Nine studies (40.9%) included secondary dependent variables involving 14 additional measurements related to math problem solving. Three studies (13.6%) also looked at students' ability to discriminate among different problem types independently (i.e., change addition or change subtraction). Two studies (9.1%) conducted pre-/post-measures of participants general problem-solving performance on a standardized math achievement assessment. Four studies (18.2%) measured students' use of the targeted strategies or diagrams when solving the problem, and another three studies (13.6%) measured student's accuracy in diagram formation, problem representation, or use of concrete manipulatives. Two studies (9.1%) implemented a measure related to knowledge of cognitive and metacognitive strategies taught during the intervention.

Table 4

Secondary Dependent Variable, Social Validity, and Generalization/Maintenance Measures

Study	Secondary dependent variable(s)	Social validity measure	Generalization/maintenance measures
Bouck & Long (2020)	None	Interviews with participants	G: Use of diagram when presented with same problem type but without given picture AND accuracy in solving a real-world representation with access to a schema M: Weekly for 4 weeks after end of intervention
Bouck et al. (2021)	None	Interviews with participants and teachers	G: Use of diagram when presented with same problem type but without given picture AND accuracy in solving similar, but different problem type with access to a schema M: 2 weeks after end of intervention
Bouck et al. (2017)	None	Interviews with participants and teachers	G: None M: 2 probes 2 weeks after end of intervention
Brawand et al. (2020)	None	Participant survey	G: None M: 4 weeks after end of intervention
Browder et al. (2012)	None	Adapted intervention rating profile completed by teacher	G: Similar problem types M: Every 2-3 weeks after intervention for the first 3 phases

Table 4 (continued)

Study	Secondary dependent variable(s)	Social validity measure	Generalization/maintenance measures
Browder et al. (2018)	Discrimination of problem types	Interviews with teachers	G: Problems presented on SMART board with virtual manipulatives AND video problems using real-world scenarios M: 1 week after end of intervention
Freeman-Green et al. (2015)	Strategy use and pre/post measures of strategy knowledge and a standardized mathematical assessment (WJ-III)	Student questionnaire	G: Different problem topics in both baseline and post-intervention M: 2 weeks and 6 weeks after intervention
Jitendra et al. (2002)	Strategy usage	Questionnaire for participants and teacher	G: Novel word problems with similar and different types than those included in instruction M: Intermittently for remainder of study after end of intervention phase with first 3 participants
Ley Davis (2016)	Discrimination of problem type	Pre/post survey for teachers; Pre/post questionnaire for tutee and tutor participants measuring perceptions and attitudes toward one another	G: Performance with an unfamiliar peer tutor M: Intermittently for remainder of study after end of intervention phase with minimum of 1 probe per participant. Alternated familiar and unfamiliar peer tutor
Maccini & Ruhl (2000)	Strategy use and problem representation	Questionnaire for participants	G: Same problem type, but presented through different story line structure AND more complex problems M: Posttest up to 6 weeks after end of intervention

Table 4 (continued)

Study	Secondary dependent variable(s)	Social validity measure	Generalization/maintenance measures
Na (2009)	None	Questionnaire for participants	G: Different problem type M: 2 probes 2 weeks after the end of intervention
Root, Cox, Davis, & Hammons (2020)	None	Open-ended questionnaire	G: How problem was presented M: None
Root, Cox, Gilley, & Wade (2020)	None	Interviews with participants and teacher	G: No graphic organizer M: None
Root, Cox, Saunders, & Gilley (2020)	None	TOMA-3 Attitude Toward Math pre/post intervention and interview with participants	G: Percent of change problem not involving tip or sale M: 3 probes after each phase of the intervention before beginning next phase
Root et al. (2018)	Discrimination of problem type	None	G: Different problem type M: 3 probes after each phase of the intervention before beginning next phase
Root et al. (2017)	None	Interview with participants	G: Use of different electronic touch-based device M: Approximately weekly after the end of the intervention through the end of the study
Saunders et al. (2018)	None	Teacher and participant survey	G: None M: Weekly probes through the end of the school year/study

Table 4 (continued)

Study	Secondary dependent variable(s)	Social validity measure	Generalization/maintenance measures
Schaefer Whitby (2013)	None	Intervention Rating Profile with teachers and Mathematical Problem Solving Assessment – Short Form with participants	G: Weekly probe completed by all peers in general education classroom averaged as a peer comparison M: 3 probes 4 weeks after end of intervention
Scheuermann et al. (2009)	Concrete manipulations, standardized math assessment (KeyMath-R)	None	G: Different problem types and presented in different word format M: 2 probes 11 weeks after end of intervention
Sheriff & Boon (2014)	None	Open-ended survey	G: None M: Minimum of 3 probes per participant; no time frame detailed
Shin & Bryant (2017)	Cognitive and metacognitive strategy questionnaire	Participant questionnaire	G: None M: None
van Garderen (2007)	Diagram use, diagram form	Participant questionnaire	G: Use of diagrams when presented with problem without task list M: 1 and 3 weeks after intervention

Social Validity. Twenty studies (90.9%) included at least one measure of social validity. Of those 20 studies, seven (35%) included structured or semi-structured open-ended interviews with participants. Six social validity measures consisted of conducting interviews with student participants, and four social validity measures consisted of conducting interviews with teachers. Of the 20 studies that included measures, 14 studies (70.0%) included a survey or questionnaire related to usability, satisfaction, and perceptions of the intervention. Eleven (55.0%) studies included social validity surveys or questionnaires receiving feedback from student participants, and 5 (25.0 %) studies included surveys or questionnaires receiving feedback from teachers involved with the study or its participants.

Generalization and Maintenance Measures. Of the 22 studies, 17 (77.3%) included a measure of generalization. Ten (58.8%) studies measured students' ability to generalize to different problem types or different problem presentations than those that were directly taught to them. Four studies (23.5%) measured students' use of diagrams and accuracy in problem solving when they were not provided the same graphic organizers and supports as during the intervention sessions. Two of the 17 studies (11.8%) measured how students performed when provided a different type of electronic device or support as compared to the devices used in the primary intervention sessions. One study generalized participants' performance to the performance to their peers in the general education class, and one study measured how participants performed with a peer tutor that was different than the one assigned to them. There were19 studies (86.4%) that included a maintenance measure. In those studies, 1-7 data probes were administered ranging from as early as an immediate session following the last intervention session to 11 weeks from when the last intervention session occurred. Most studies (n = 15; 78.9%) included probes between 1-4 weeks after the end of the intervention.

Self-Regulated Learning Components

The field of self-regulated learning (SRL) research outlines three generally accepted cyclical phases in the process of self-regulation: forethought, performance, and self-reflection. Within these phases, there are specific subprocesses that have been operationally defined over the years in a variety of ways that have allowed for studies to measure their impact on an individual's performance on a given task. All studies in the sample were screened for self-regulated learning intervention components including cognitive strategy, goal setting, planning, self-monitoring, self-control, and self-evaluation. The last five are considered specific metacognitive skills frequently involved in a robust self-regulated learning process (Dent & Koenka, 2016). Cognitive strategies were defined as explicit or implicit processes used during an academic task that are intertwined with aspects of self-regulation leading to accomplishing the task (Dent & Koenka, 2016). Goal setting was defined as explicitly setting a goal for a future performance or session, oftentimes based on previous performances. Planning was defined as explicit activation of previous metacognitive skills independently to outline next steps. Self-monitoring was defined as tracking progress through the completion of the task and learning process. Self-control was defined as explicitly changing strategies to perform at a higher level in future attempts. Self-evaluation was defined as explicitly comparing that day's performance to performance on previous sessions. Studies were screened using these terms and definitions to determine the presence of self-regulated learning strategies within the intervention package.

The different approaches for how these aspects of self-regulated learning are measured in studies can be grouped into two distinct categories: offline and online measures. Offline measures are completed separate from the task itself through questionnaires or interviews related to strategy knowledge, strategy usage, or self-efficacy related to the academic skill at hand.

Online measures involve observable and measurable evidence of students using cognitive strategies as they complete tasks. This could include think-aloud protocols to hear students' thought process in real time or the analysis of work on paper such as underlining or working through a task list. Studies were coded for evidence of offline or online measures that were present in the study, regardless of whether the evidence was coded or analyzed.

For the content data collected in this meta-analysis, seven studies (31.8%) were cross-checked by another researcher to gather interrater reliability (IRR) for content coding accuracy on sample size, grades of participants, diagnoses/eligibility category, race/ethnicity reported, group size, math skill targeted, curriculum-based material, interventionist, setting, components of SRL included, and offline/online SRL measures present. The mean IRR by component was 93.51% and the mean IRR by article was 93.51%. Table 5 presents components of SRL used during the intervention phases across all included 22 studies. Fourteen studies (63.6%) included an explicit component of self-monitoring. Two studies (9.1%) included an explicit component for goal setting. Four studies (18.2%) included an explicit component for self-evaluation. None of the studies included explicit components for planning or self-control as defined above. Only three studies (13.6%) included an offline measure of skills related to SRL. Two of these were strategy knowledge pre- and post-assessments and one involved a strategy questionnaire. All the studies incorporated evidence of online SRL traces including marking up graphic organizers or diagrams or marking through a task analysis checklist, but only eight (36.4%) of them analyzed the data available to them as such. Four studies (18.2%) included think-aloud protocols as an online measure of students' SRL throughout the problem-solving process.

Question 2. Quality of Studies

In 2014, the Council for Exceptional Children (CEC) published a set of standards to

Table 5

Components of Self-Regulated Learning (SRL) Included in Interventions

Study	Component(s) of SRL	Online and/or offline measure
Bouck & Long (2020)	CS	Online: completed graphic organizer
Bouck et al. (2021)	CS	Online: completed diagram
Bouck et al. (2017)	CS	Online: traces of manipulatives and visual representations
Brawand et al. (2020)	CS, SM	Online: schematic diagram markups and checklist of steps
Browder et al. (2012)	CS	Online: completed graphic organizer
Browder et al. (2018)	CS, SM	Online: completed graphic organizer and checklist of cognitive/metacognitive steps
Freeman-Green et al. (2015)	CS, SM	Online: think aloud questions Offline: strategy use pre/post measure
Jitendra et al. (2002)	CS, SM	Online: traces of strategy use and marking of word problem Offline: strategy questionnaire
Ley Davis (2016)	CS, SM	Online: completed graphic organizer and checklist of steps
Maccini & Ruhl (2000)	CS, SM	Online: think aloud comments analyzed and coded for use of sub-steps of strategies
Na (2009)	CS, SM	Online: completed schema diagram and checklist of steps
Root, Cox, Davis, & Hammons (2020)	CS, SM, SE	Online: completed mark ups of problems and graph of personal results
Root, Cox, Gilley, & Wade (2020)	CS	Online: completed graphic organizer
Root, Cox, Saunders, & Gilley (2020)	CS, GS, SM, SE	Online: written goals set for next session, checklist of steps/marking word problem, and graph of session progress
Root et al. (2018)	CS	Online: completed graphic organizer
Root et al. (2017)	CS, SM	Online: completed graphic organizer and checklist of steps

Table 5 (continued)

Study	Component(s) of SRL	Online and/or offline measure
Saunders et al. (2018)	CS, SM	Online: task checklist of steps
Schaefer Whitby (2013)	CS, SM, SE	Online: say/ask specific think-aloud prompts during task components and graph of session progress Offline: strategy knowledge pre-/post-questionnaire
Scheuermann et al. (2009)	CS	Online: think aloud comments noted
Sheriff & Boon (2014)	CS	Online: printed traces of completed graphic organizer from computer
Shin & Bryant (2017)	CS, GS, SM, SE	Online: set goals for next session, checklist of task steps and marking of word problem, and graph of session progress
van Garderen (2007)	CS, SM	Online: completed drawn diagram

Note. CS = cognitive strategy; GS = goal setting; SM = self-monitoring; GS = goal setting; SE= self-evaluation.

measure and compare the strength of study design when investigating evidence-based practices (CEC, 2014; Royer et al., 2017). The CEC's standards include 8 quality indicators: context and setting, participants, intervention agent, description of practice, implementation fidelity, internal validity, outcome measures/dependent variables, and data analysis. Studies received both absolute and weighted ratings for quality indicators met. When using absolute coding, studies earned a "1" for each quality indicator in which they met all components included. If there was a single component that was not met, the study earned a "0" for that component when implementing absolute coding. Weighted coding allowed for minor lapses in study design or reporting by providing weighted scores for each indicator (i.e., if 3 of 5 components were met, that indicator would be noted with a weighted score of 60%). When determining studies that could be considered methodologically sound with absolute coding, studies had to meet all components of all 8 quality indicators. When determining studies that could be considered methodologically sound with weighted coding, studies had to meet a threshold of at least 80% of the components met (overall rating of 6.40 or higher quality indicators met).

All studies were coded by two researchers to obtain inter-rater reliability (IRR) measures for each component of each study and confirm measurements of methodology quality using both absolute and weighted measures. Table 6 displays the absolute rating, weighted rating, and mean

Table 6

CEC's 2014 Quality Indicators Met by Study

Study	Quality indicators met by coding method		IRR
	Absolute	Weighted	
Bouck & Long (2020)	8.0	8.0	100%
Bouck et al. (2021)	7.0	7.67	95.45%
Bouck et al. (2017)	8.0	8.0	95.45%
Brawand et al. (2020)	8.0	8.0	95.45%
Browder et al. (2012)	7.0	7.83	95.45%
Browder et al. (2018)	7.0	7.80	95.45%
Freeman-Green et al. (2015)	7.0	7.67	81.82%
Jitendra et al. (2002)	6.0	7.63	77.27%
Ley Davis (2016)	8.0	8.0	100%
Maccini & Ruhl (2000)	6.0	7.47	88.36%
Na (2009)	7.0	7.67	77.27%
Root, Cox, Davis, & Hammons (2020)	8.0	8.0	100%
Root, Cox, Gilley, & Wade (2020)	8.0	8.0	100%
Root, Cox, Saunders, & Gilley (2020)	8.0	8.0	90.91%
Root et al. (2018)	8.0	8.0	100%
Root et al. (2017)	8.0	8.0	100%
Saunders et al. (2018)	8.0	8.0	100%
Schaefer Whitby (2013)	7.0	7.67	81.82%
Scheuermann et al. (2009)	5.0	6.60	77.27%
Sheriff & Boon (2014)	8.0	8.0	100%
Shin & Bryant (2017)	8.0	8.0	100%
van Garderen (2007)	5.0	6.47	100%

interrater reliability (IRR) measure for each study. The mean IRR by article was 93.18%. The mean IRR by quality indicator component was 93.14%. Twelve studies (54.5%) met all eight quality indicators using absolute coding (and therefore had scores of 8.0 on weighted coding as well). The other 10 studies (45.5%) that did not meet criteria under absolute coding did meet the threshold for methodology quality when using weighted coding of at least 6.40 quality indicators met. It should be noted that Eight studies (36.4%) were published before the publication of the CEC (2014) standards for measuring quality indicators of studies.

Question 3. Aggregate Effect Size

Individual Tau-U effect sizes for studies were calculated using the online calculator at https://www.singlecaseresearch.org (Vannest et al., 2016). The calculator incorporates calculations of the Mann-Whitney U statistic when calculating an effect size without need of baseline trend correction. When trend correction is necessary, the Kendall's S statistic is calculated to adjust for any positive trend that may exist in the baseline data already (Brossart et al., 2018; Peltier et al., 2018).

Six studies (27.3%) included measures of multiple problem types. All the data was included if each problem type was algebra-based (i.e., group change with addition/subtraction and multiplicative comparison). Five of the six studies met those criteria and included multiple AB comparisons per participant as additional repetitions toward the overall effectiveness of the given intervention package. One study, Browder et. al (2012), conducted a multiple baseline across condition (curriculum units) with each of their four participants. The algebra unit included word problems with one-step equations involving any of the four operations, so each participant's results from this condition alone were included in the overall analysis.

Weighted average Tau-U effect sizes and 90% confidence intervals were then calculated through a combination of statistical analysis including number of participants, data points in baseline and intervention conditions, baseline trend corrections if needed, and AB comparisons within a given study (see Table 7). Nineteen studies (86.4%) returned a large effect size (Tau-U = 0.8 or larger) and the other three studies (13.6%) returned a moderate effect size (Tau-U = 0.5-0.8). The 90% confidence intervals for studies with large effect sizes ranged as low as 0.518, and the moderate effect sizes ranged as low as 0.497. Even when considering confidence intervals,

Table 7

Tau-U Effect Sizes by Study

Study	n	Data points	Phase contrasts	Tau-U	90% CI
Bouck & Long (2020)	3	33	3	1.0	[0.638, 1.000]
Bouck et al. (2021)	4	39	4	1.0	[0.674, 1.000]
Bouck et al. (2017)	4	67	4	1.0	[0.752, 1.000]
Brawand et al. (2020)	9	52	3	0.864	[0.585, 1.000]
Browder et al. (2012)	4	70	4	0.766	[0.497, 0.923]
Browder et al. (2018)	6	201	18	0.996	[0.830, 1.000]
Freeman-Green et al. (2015)	6	66	6	0.888	[0.642, 1.000]
Jitendra et al. (2002)	4	60	4	0.949	[0.668, 1.000]
Ley Davis (2016)	4	144	8	1.0	[0.757, 1.000]
Maccini & Ruhl (2000)	3	30	3	1.069	[0.692, 1.000]
Na (2009)	4	24	4	0.972	[0.553, 1.000]
Root, Cox, Davis, & Hammons (2020)	3	51	3	1.0	[0.714, 1.000]
Root, Cox, Gilley, & Wade (2020)	3	69	3	0.910	[0.673, 1.000]
Root, Cox, Saunders, & Gilley (2020)	3	150	6	1.0	[0.740, 1.000]
Root et al. (2018)	2	52	4	1.0	[0.695, 1.000]
Root et al. (2017)	3	88	6	1.027	[0.806, 1.000]
Saunders et al. (2018)	3	105	3	0.982	[0.759, 1.000]
Schaefer Whitby (2013)	3	42	3	0.953	[0.627, 1.000]
Scheuermann et al. (2009)	14	106	14	0.768	[0.559, 0.978]
Sheriff & Boon (2014)	3	26	3	0.973	[0.552, 1.000]
Shin & Bryant (2017)	3	38	3	0.766	[0.400, 1.000]
van Garderen (2007)	3	43	3	0.858	[0.518, 1.000]
Aggregated mean	94	1556	112	0.930	[0.750, 1.000]

Note. n = number of participants; AB = number of AB contrasts; CI = confidence interval.

all studies had at least a moderate positive effect on participants' independent mathematic problem-solving performance. Two studies returned a Tau-U effect size greater than 1 (Maccini & Ruhl, 2000; Root et al., 2017). Vannest et al. (2016) do briefly address the possibility of this occurrence on the web-based application, but the implications of this anomaly will be discussed in depth within the discussion.

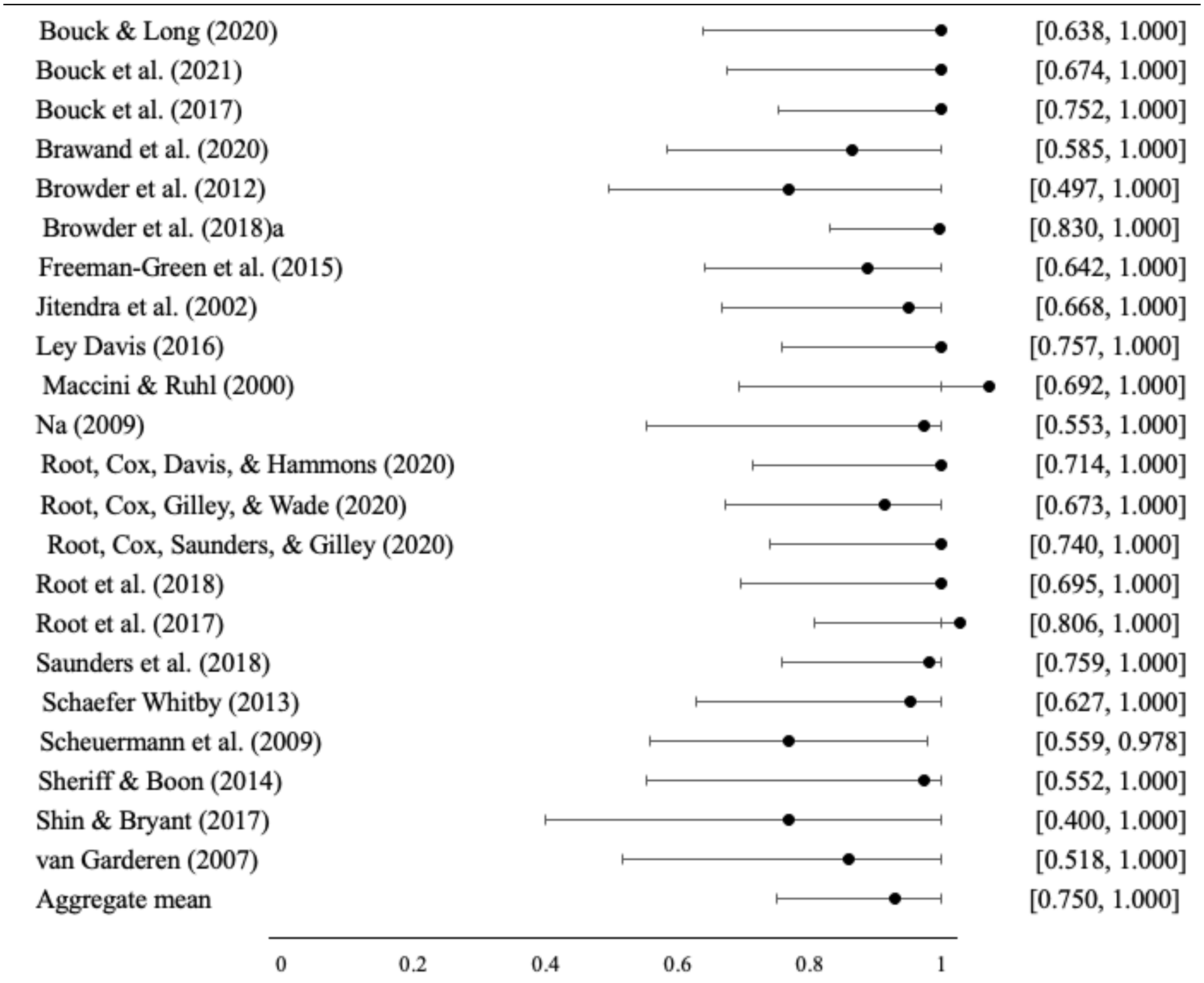
The aggregate mean effect size was calculated using the Jamovi statistical program's meta-analysis module. The overall Tau-U aggregate mean effect size was 0.93 (90% CI = [0.750, 1.000]), demonstrating that increases observed in independent math problem solving performance across the full sample may very likely be due to the implementation of the interventions involved (see Figure 3). More specifically 93% of the intervention data points that improved above baseline are likely due to the interventions.

To analyze the robustness of these results, tests were conducted for heterogeneity, publication bias, and sensitivity analysis. Design-related heterogeneity was addressed during the study design phase by ensuring that all studies involved included multiple baseline designs and involved individuals with disabilities in middle school. It is noted that some heterogeneity exists in this manner due to the wider design of all disability categories instead of a focus on a single disability category (i.e., students with learning disabilities, intellectual disability, or autism). Next a statistical measure was computed for between-study heterogeneity. Thompson's I^2 was used to confirm extremely low heterogeneity between studies in this sample ($I^2 = 0\%$), supporting the credibility of the aggregate effect size results.

Publication bias was analyzed using the Rosenthal approach for a Fail-safe N calculation. This statistical measure determines how many additional studies with a negative effect would need to be added to bring the overall statistical power into a non-significant range. Therefore,

Figure 3

Forest Plot for Study Effects



high Fail-safe N values indicate a stronger probability that publication bias did not impact the studies available and included in the results of the current meta-analysis. This meta-analysis demonstrated that publication bias likely did not impact the overall aggregate mean effect size (fail-safe N = 845.00, $p < 0.001$).

A sensitivity analysis was conducted by rerunning the aggregate mean effect size with only studies that met all 8 CEC quality indicators using absolute coding measurements ($n = 12$). When only including studies that met the higher standard of methodological quality analysis and clarity in reporting, the aggregate mean effect size was 0.96 (90% CI = [0.73, 1.00]). This analysis also included a statistically acceptable homogeneous data set ($I^2 = 0\%$) and adequate protection from potential publication bias impacting the overall results (fail-safe N = 271.00, $p < 0.001$). Based on these assessments of heterogeneity, publication bias, and sensitivity analysis, the aggregate mean effect size results should be deemed robust and suitable for further interpretation as otherwise deemed appropriate.

Question 4. Impact of Disability on Effect Size and Other Potential Moderators

After reviewing results of the overall aggregate effect size for studies included in the meta-analysis, the next process was to determine if effect sizes had any significant variations among groups of participants based on disability category or status. Analysis was run on a dataset including all 94 participants, their demographic data, and the study information in which they participated (see Table 8). Secondary moderator analyses were also completed to analyze the potential impact of interventionist, group setting, intervention type, and problem type on overall effect sizes. The SPSS meta-analysis statistical program was used for moderator analysis computations. The participant dataset was used for all moderator analyses versus the group study dataset to provide a more conservative aggregate effect size and a tighter confidence interval due to a more accurate increased sample size. For all moderator statistics computed, each analysis passed tests for heterogeneity (all had $I^2 = 0\%$). These analyses demonstrated high effect sizes across all categories with results consistent with the results of individual studies and of the overall aggregate mean.

Table 8

Moderator Analysis

Group	n	Data points	Phase contrasts	Tau-U	90% CI
Disability Category					
Learning Disability	45	517	45	0.899	[0.808, 1.000]
Intellectual Disability	33	860	54	0.962	[0.857, 1.000]
Autism	13	318	16	0.923	[0.772, 1.000]
ADHD + Other Health Impairment + Auditory Processing Disorder	3	38	3	0.814	[0.489, 1.000]
Interventionist					
Researcher	70	1175	78	0.927	[0.858, 1.000]
Special Education Teacher	17	377	29	0.907	[0.754, 1.000]
Peers [a]	4	143	8	1.000	[0.830, 1.000]
Technology/Computer Program [a]	3	38	3	0.766	[0.400, 1.000]
Group Setting					
Individual	52	1088	64	0.963	[0.883, 1.000]
Small group	42	645	54	0.861	[0.765, 1.000]
Intervention Type					
General problem-solving variations	27	302	27	0.830	[0.704, 1.000]
CRA variations	10	166	10	0.974	[0.820, 1.000]
SBI	21	383	25	0.914	[0.790, 1.000]
MSBI	17	598	37	1.000	[0.832, 1.000]
Diagram-based approaches (not SBI/MSBI)	10	115	10	0.950	[0.756, 1.000]
Technology-based approaches	9	169	9	0.935	[0.764, 1.000]
Problem type					
Percent of change (ratios/fractions/cost)	16	311	19	0.956	[0.800, 1.000]
Multiplication/Division (including change problem type)	11	153	11	0.933	[0.768, 1.000]
Addition/Subtraction (including change problem type)	19	539	28	1.000	[0.876, 1.000]
One-variable word problem	26	378	38	0.863	[0.717, 1.000]
Proportional reasoning [a]	9	156	9	0.864	[0.585, 1.000]
Unclear	13	196	13	0.921	[0.860, 1.000]

Note. [a] Moderator category only includes data from a single study.

Risk of Bias Assessments

Looking at potential risk for bias in studies included within a meta-analysis allows for better conclusions to be drawn about the measured aggregate mean effect size and the conclusions drawn from that information. To address risk of bias, each study was coded using a checklist based on Reichow et al.'s (2018) Single Case Design Risk of Bias (SCD RoB) tool. Reichow et al. designed the SCD RoB tool using the Cochrane Risk of Bias tool and framework that is widely used in meta-analyses involving randomized controlled trials. The SCD RoB incorporates standards and language clarification of single-case design that is more unclear in Cochrane's original tool designed for randomized controlled trials. The SCD RoB tool includes eight domains across selection, performance, and detection biases: sequence generation, participant selection, blinding of participants and personnel, procedural fidelity, blinding of outcome assessment, selective outcome reporting, dependent variable reliability, and data sampling.

Articles were reviewed and coded with a low, unclear, or high risk of bias based on the criteria outlined by Reichow et al.'s (2018) framework (see Figure 4). Study ratings were then combined to create overall potential for bias across each domain for all studies included in the meta-analysis (see Figure 5). For sequence generation (selection bias), 11 studies (50.0%) were rated as low risk of bias and the other 11 studies (50.0%) were rated as unclear risk of bias. This measure looked at methods studies reported to decide on an order for introducing participants to interventions or different conditions. Many studies either reported randomization techniques utilized for introducing participants in a certain order or discussed the decision-making process that occurred to determine the order that was implemented. Others had clearer decisions based on differences in baseline trends among participants. Many studies that received unclear ratings

Figure 4

Summary of Risk of Bias by Study

Study	Selection Bias		Performance Bias		Detection Bias			
	SG	PS	BPP	PF	BOA	SOR	DVR	DS
Bouck & Long (2020)	?	+	-	+	-	?	+	+
Bouck et al. (2021)	?	?	?	?	-	+	+	+
Bouck et al. (2017)	?	+	+	?	-	+	+	+
Brawand et al. (2020)	+	?	?	+	?	+	+	+
Browder et al. (2012)	+	?	?	?	+	?	?	+
Browder et al. (2018)	?	+	?	?	+	+	?	+
Freeman-Green et al. (2015)	+	+	?	?	?	+	+	+
Jitendra et al. (2002)	?	+	+	?	+	+	+	+
Ley Davis (2016)	?	+	?	+	+	+	+	+
Maccini & Ruhl (2000)	?	+	-	+	?	+	?	+
Na (2009)	+	+	+	?	?	?	?	+
Root, Cox, Davis, & Hammons (2020)	?	+	?	+	?	+	+	+
Root, Cox, Gilley, & Wade (2020)	+	?	?	+	+	+	+	+
Root, Cox, Saunders, & Gilley (2020)	+	+	?	+	?	+	+	+
Root et al. (2018)	?	+	?	+	?	+	+	+
Root et al. (2017)	+	+	?	+	?	+	+	+
Saunders et al. (2018)	+	+	?	+	-	+	?	+
Schaefer Whitby (2013)	+	+	?	?	+	+	?	+
Scheuermann et al. (2009)	+	+	+	-	?	-	+	+
Sheriff & Boon (2014)	?	-	?	+	+	?	+	+
Shin & Bryant (2017)	+	+	?	?	+	+	+	+
van Garderen (2007)	?	+	+	-	?	+	+	?

Note. SG = Sequence generation; PS = Participant selection; BPP = Blinding of participants and personnel; PF = Procedural fidelity; BOA = Blinding of outcome assessment; SOR = Selective outcome reporting; DVR = Dependent variable reliability; DS = Data sampling; $\boxed{+}$ = Low risk of bias; $\boxed{?}$ = Unclear risk of bias; $\boxed{-}$ = High risk of bias. Adapted from "Development and

Figure 4 (continued)

applications of the single-case design risk of bias tool for evaluating single-case design research study reports," by B. Reichow, E. E. Barton, & D. M. Maggin, 2018, 79, 53–64 (https://doi.org/10.1016/j.ridd.2018.05.008). Copyright 2018 by Elsevier Ltd. Reprinted with permission; permission conveyed through Copyright Clearance Center, Inc.

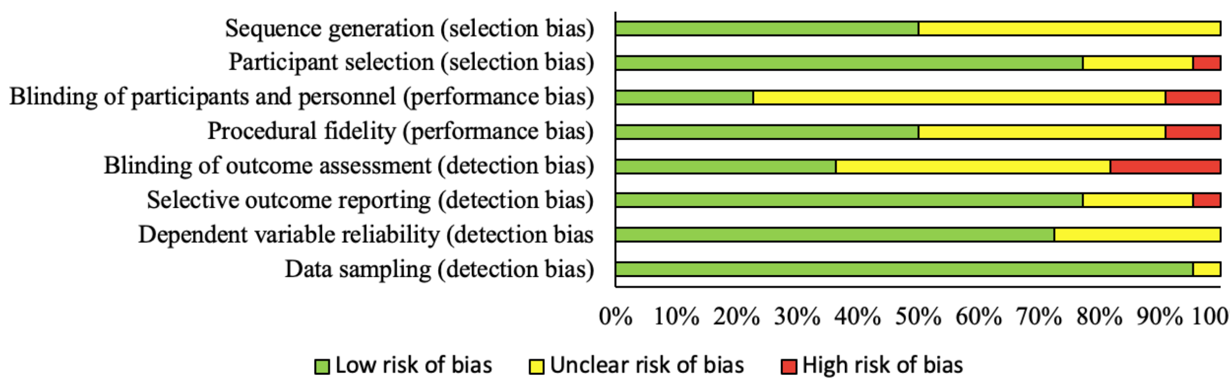
included at least two participants with identical baseline data and did not report a decision process related to initial sequence generation.

For participant selection (selection bias), 17 studies (77.3%) were rated as a low risk of bias, four studies (18.2%) were rated as an unclear risk of bias, and one study (4.5%) was rated as a high risk of bias. This measure analyzed the criteria and process involved in determining adequate participants to recruit for study participation. Many studies specifically screened potential participants for performance on the proposed dependent variable to confirm lacking skills in this area on top of reporting additional inclusion criteria. Studies that were rated as unclear did not report if the source of information for participants' performance on the specific target skill and/or if it was observed prior to final inclusion in the study or not. One study did not report detailed inclusion methods, how individuals performed on the target skill prior to inclusion, or any information about how it was determined that those individuals were appropriate for the intent of the study.

For blinding of participants and personnel (performance bias), five studies (22.7%) were rated as a low risk of bias, 15 studies (68.2%) were rated as an unclear risk of bias, and two studies (9.1%) was rated as a high risk of bias. This area of bias analyzed methods utilized to

Figure 5

Summary of Risk of Bias by Domain



Note. Adapted from "Development and applications of the single-case design risk of bias tool for evaluating single-case design research study reports," by B. Reichow, E. E. Barton, & D. M. Maggin, 2018, 79, 53–64 (https://doi.org/10.1016/j.ridd.2018.05.008). Copyright 2018 by Elsevier Ltd. Reprinted with permission; permission conveyed through Copyright Clearance Center, Inc.

increase unawareness of intervention implementation and conditions for members of the research team that were making decisions regarding condition changes. Most studies generally reported specific parameters for when to change conditions (mastery criteria standards), but it was unclear based on wording if these were set prior to initial implementation.

For procedural fidelity (performance bias), 11 studies (50.0%) were rated as a low risk of bias, nine studies (40.9%) were rated as an unclear risk of bias, and two studies (9.1%) was rated as a high risk of bias. This measure evaluated the details reported for training external interventionists and for tracking interventionists implementing the steps of each condition as designed. Studies rated with a low risk of bias reported procedural fidelity ratings for each condition and training session if applicable, including at least 20% of probes in each and maintained 80% or greater accuracy in implementation. Studies rated with an unclear risk of bias only included procedural fidelity measures for the intervention phase and did not report information about baseline or other additional conditions. Studies rated with a high risk of bias did not include procedural fidelity ratings of any sort.

For blinding of outcome assessment (detection bias), eight studies (36.4%) were rated as a low risk of bias, 10 studies (45.5%) were rated as an unclear risk of bias, and four studies (18.2%) was rated as a high risk of bias. This measure critiqued the reported methods in place to minimize or eliminate awareness of research team members for when the intervention was implemented to specific participants (i.e., which probes were for intervention or baseline). As is common in many acceptable single-case design studies currently, the primary interventionist also served as the primary data collector in many studies which would not provide protection against this type of bias. Studies rated with a low risk of bias reported that both data collectors for interrater coding were different than the interventionist or that the secondary rater coded 100% of probes. Studies rated with a high risk of bias reported that both individuals who were involved to measuring student performance and decisions made about interventions were trained interventionists involved in the study.

For selective outcome reporting (detection bias), 17 studies (77.3%) were rated as a low risk of bias, four studies (18.2%) were rated as an unclear risk of bias, and one study (4.5%) was rated as a high risk of bias. This measure analyzes the outcome data reported and if any data was missing or unreported. Studies that were rated as an unclear risk of bias involved other students when the intervention was taking place but did not report on their characteristics or their performance on the intervention. Other studies excluded students due to heightened performance near the beginning of the study, but it was unclear if this occurred before or during baseline condition. Another reported that a participant had to withdraw due to excessive absences. Studies that were rated as a high risk of bias reported multiple participants were excluded after the baseline phase due to increased variability in their data, which was not reported elsewhere in the results.

For dependent variable reliability (detection bias), 16 studies (72.7%) were rated as a low risk of bias, six studies (27.3%) were rated as an unclear risk of bias, and no studies (0.0%) was rated as a high risk of bias. This measure looked at how inter-observer and/or inter-rater agreement was completed and to what level agreement was reported. Studies that were rated as an unclear risk of bias often did not report a breakdown of sessions double-coded per condition versus an overall percentage. Studies that were rated as a high risk of bias included statistics for inter-observer agreement per condition and had levels of agreement at least 80% or higher.

For data sampling (detection bias), 21 studies (95.5%) were rated as a low risk of bias, one study (4.5%) were rated as an unclear risk of bias, and no studies (0.0%) was rated as a high risk of bias. This measure analyzed if the data available and the trends produced met adequate standards for functional relationships to be established. Most studies were rated as a low risk of bias and included appropriate amounts of data points in each condition per participant and demonstrated adequate trends at condition changes to support prediction, verification, and replication. Studies that were rated as an unclear risk of bias included enough data points, but there were questionable trends in baseline regarding condition change decisions and there was only one baseline datapoint that overlapped for all three participants.

Certainty of Evidence

Based on these results, sequence generation, participant selection, selective outcome reporting, dependent variable reliability, and data sampling are reliable for low risks of bias within this meta-analysis. Blinding of participants and personnel, blinding of outcome assessment, and procedural fidelity are less reliable with overall unclear risks of bias. Common aspects of acceptable single-case research today do not often require complete blinding of participants and personnel or blinding of outcome assessment as directly as stated in the risk of bias tool framework utilized in this coding. These areas are often kept in check to some extent through additional research team members involvement in different aspects of the study to ensure reliability regarding performance and detection of change. Altogether it is unlikely that any bias present in the design and implementation of the studies strongly impacted the overall results of this analysis.

CHAPTER 4

Discussion

This chapter consists of the interpretation and analysis of the meta-analytic results according to each research question. This chapter also includes a discussion on the limitations of this study and directions for future research. Finally, implications for future research and practice based on the results and analysis will be presented.

Relationship of Review Findings to Existing Literature

The aggregate effect size of math problem-solving interventions with components of self-regulated learning for middle school students with disabilities presented in this study is consistent with similar studies that exist in the literature. Mathematic word problem solving had been found effective across a synthesis of both single-case design studies (ES = 0.90) and group design studies (ES = 0.95) when targeting students with math-related disabilities (Zheng et al., 2012). Additionally, Peltier et al. (2018) established an aggregate Tau-U effect size of 0.88 when analyzing 16 single-case studies that implemented schema-based instruction across all grades, which is commensurate with the effect for SBI that this analysis found for SBI specifically with middle school students. One study included by Peltier et al. was included in this analysis as well (e.g., Jitendra et al., 2002).

In relation to math problem solving instructional components, this analysis supports the conclusions that making accurate graphs or diagrams to represent a problem are effective strategies for students with disabilities. Krawec (2014) had determined that these strategies were more effective with students with disabilities than with their average achieving counterparts. The instruction of visual representation was present in all studies within this analysis to some extent, demonstrating that they are indeed effective with middle school students with disabilities.

Regarding the self-regulated learning (SRL) literature, this meta-analysis supports multiple previously noted concerns about SRL interventions. First, interventions that had both cognitive and metacognitive components were effective with students with disabilities. Babakhani (2011) found consistent results when assessing verbal math problem-solving skills with elementary students. Understandably though, these strong effect sizes that have been found across multiple studies and meta-analyses at a direct intervention level may not generalize all the way to standardized achievement tests and other far-generalized measures (Dent & Koenka, 2016). A noted concern in the SRL literature involved not addressing all cyclical components of SRL within an intervention package (Reddy et al., 2018). Reddy et al. found through a systematic review that most interventions focused solely on aspects of SRL within the performance phase and neglected to close the loop with forethought and reflection phase components (e.g., goal setting and self-reflection). Similarly, the SRL components included in the studies of this meta-analysis primarily addressed the performance phase with specific cognitive strategies or self-monitoring strategies. Only two studies included both a version of goal setting and self-evaluation (e.g., Root, Cox, Saunders, & Gilley, 2020; Shin & Bryant, 2017).

Dignath and Büttner (2008) found that teaching motivation strategies alongside metacognitive components was more effective with secondary students than a sole focus on a given cognitive strategy. Over half of included studies in this analysis (n = 14; 63.6%) incorporated multiple SRL components beyond a single cognitive strategy, but none incorporated motivation as a key aspect of the intervention, related to either academic skills or SRL. The results of this meta-analysis compliment and strengthen the existing literature and support the use of math problem-solving interventions with SRL components specifically for middle school students with disabilities.

Limitations of Review Process

There are several limitations of this review and analysis. First and foremost, as the case with many systematic reviews and meta-analysis studies, this review was limited to only articles that were available in English, which could have likely excluded some studies that would have otherwise met the inclusion criteria. Another general limitation is that most studies published in peer-reviewed journals have positive results and do not include studies that contain unfaorable outcomes. This may lead to inflated interpreations about the effectiveness of an intervention. There is a possibility that the researcher may have missed articles that did not appear through the data-based and ancestral searches. This is more likely if the titles and abstracts did not contain content needed to meet the inclusion criteria and therefore were passed by for inclusion in the next round of review. In these cases, the researcher and independent raters would not have fully read the articles to determine if they did in fact meet inclusion criteria. Additionally, this study was limited by only including single-case design studies and by excluding any studies published prior to 2000. In both instances, there were likely important studies that could have met inclusion criteria that would have added to the strength and scope of the current study. For example, the Solve It technique, only utilized by one single-case design study included in this review, is a well-researched intervention. Single-case design studies involving this intervention were found to have been completed prior to 2000 by Montague and colleagues. Additionally, more recent research involving this technique implemented the intervention through group designs within large-scale randomized controlled trials. Because of the current review parameters, these studies were not included in the present analysis and therefore an important limitation to note.

Tau-*U* is one of a handful of options available to measure, compare, and combine effect sizes statistically within single-case design studies. It has been used often as an effect size measure for single-case design (SCD) studies (Lo et al., 2015; Root et al., 2018; Ross & Sabey, 2015; Shin & Bryant, 2015; Shin & Bryant, 2017), and furthermore in meta-analyses involving SCD studies (Long et al., 2019; Peltier et al., 2018). Tau-*U* is considered particularly useful due to its ability to correct for non-conforming baseline trends (Lee & Cherney, 2018) and still may be the most relevant measure for within-case comparisons (Kingbeil, et al., 2019). As with any statistical analyses though, there are limitations to Tau-*U*'s calculation and interpretation. A leading limitation of Tau-*U* is that it specifically measures within-participant variation of data, and therefore cannot be broadly compared to effect sizes gathered from group design studies (Maggin et al., 2017). Perhaps of greater concern, the underlying calculations of Tau-*U* using the online calculator at https://singlecaseresearch.org at times inflates results and computes Tau-*U* measures that are outside the bounds of -1 and +1. This is most significant because even measures that fall within the normal limits of -1 and +1 could be inflated, leading to inaccurate interpretations of overall results (Brossart et al., 2018; Kingbeil et al., 2019). Two of the studies in this analysis (e.g., Maccini & Ruhl, 2000; Root et al., 2017) produced Tau-*U* effect sizes greater than 1. Moreover, there may also be the possibility of inflated effect sizes for some of the other studies in this review due to large variability within the data set even in cases where there is no overlapping between baseline and intervention phases.

Another concern observed after conducting the analysis using Tau-*U* involved reported measures of Tau-*U* in studies. Most studies analyzed data using visual analysis and a simple percentage of non-overlapping data per participant. One study did provide Tau-*U* calculations of the overall effect size, calculated using the online calculator at www.singlecaseresearch.org (e.g., Root et al., 2018). Root et al. (2018) reported an overall Tau-*U* effect size of 0.87 compared to the calculated effect size of 1.0 from the procedures used in this study. This difference could be because the researchers were able to use exact data points from original data versus data extracted from graphs. It could also be from an inclusion of different data points within the online calculator. The Root et al. (2018) study involved multiple subphases and the researchers did not clearly state exactly which data points were included in the calculation of the Tau-*U* effect size, so it is possible that the reported effect size did not include all the phase data or computed information across different problem types.

As is the case with any single-case design research, sample sizes are small when conducting additional analyses with potential moderators. Many conclusions drawn from the moderator statistics are limited due to the small sample sizes or representation of a single study from within the review. There is not a significant difference in intervention effects as measured by effect sizes among disability groups.

Finally, conclusions made regarding risk of bias present among studies in the sample are limited due to the use of a description of a framework for single-case design studies without additional materials, content validity measures, or replication support. Primary documents or protocols were not available even after reaching out directly to the Reichow et al. (2018) corresponding author. Additionally, no interrater reliability (IRR) coding was completed on the risk of bias analysis. The use of IRR coding in this instance could have increased the value of the current study in relation to greater use of Reichow et al.'s (2018) SCD RoB tool.

Interpretation and Limitations of Individual Studies

Study Characteristics

Studies included in the analysis incorporated a variety of different characteristics. Most studies included three participants, which is a common standard for single-case research design (SCRD) studies using a multiple-probe or multiple-baseline designs. A minimum requirement in single case design studies is three phase replications, although more are recommended if possible (Lobo et al., 2017). Ten studies did include more participants and replications, but all fell within an average range for SCDs with no more than 15. The inclusion of additional participants strengthens the results of the overall study by further minimizing potential threats to internal and external validity. There were more male students involved in the studies than female students, which is consistent with overall patterns of students who receive special education services within the United States (Irwin et al., 2021). The Race/ethnicity of samples of participants across studies in this review was representative of overall population data (United States Census Bureau, 2021) than it was representative of the recent special education population data (Irwin et al., 2021).

Disability Representation. Although different disabilities were represented in the sample, about 50 percent of the participants were identified as having a learning disability. This finding was not surprising given that the most recent annual *Report on the Condition of Education* indicated that specific learning disability (SLD) was the category under which the greatest percentage of students qualify for special education services (Irwin et al., 2021). Students identified with and receiving special education services for an intellectual disability were the next highest category represented in this study with 33 participants (35%). Although students with intellectual disabilities make up only 6% of all students identified with disabilities

(Irwin et al., 2021), the MSBI intervention approach aims to modify successful schema-based instruction models so that it is appropriate and effective for students with moderate intellectual disabilities. Although not analyzed more specifically in the results, it should also be noted that six of the students included in the sample with intellectual disabilities had a more specific medical diagnoses of Down syndrome. Across the studies, the sample of students identified with Autism (13; 13.8%) mirrored the overall percentage of this population (11%) in the United States currently receiving special education services (Irwin et al., 2021).

Interestingly, there were no studies in this review that included middle school participants with emotional disturbance (ED). However, there have been studies and systematic reviews that have included elementary grade students with ED in studies examining math performance or self-regulation (Alter, 2012; Alter et al., 2011; Hawkins & Heflin, 2011; Popham et al., 2018; Tan, 2016). Moreover, few studies have focused on middle school students with ED's problem-solving skills specifically. Jitendra et al. (2010)'s case study found promising results from implementing SBI with two middle school students with ED, and Mulcahy and Krezmien (2009) demonstrated generally positive results when measuring middle school students with ED's math performance accuracy on geometry problems. Because of the nature of ED disabilities hindering one's self-regulation skills in a variety of ways, it would be beneficial to see more research in this area for students with ED specifically.

Setting and Interventionist. With regard to the setting and geographic locations of the studies in this review, urban, rural, and suburban geographic locations and public, private, and charter school settings were represented across multiple studies, as were. All studies that reported on the geographic region within the United States indicated that the studies took place in either the Midwest or eastern (north and south) regions of the country.

Furthermore, most interventions were conducted in private separate spaces away from the participants' normal classroom environment. This is also common for single-case design studies in special education research when researchers are the interventionists. In contrast, all four interventions that were conducted in the special education classroom setting were conducted by the intervention specialist/special education teacher who had been trained by the researchers. It is more work to ensure adequate levels of fidelity across implementation of interventions when using a direct training or train-the-trainer model in research, but it has the potential for greater impact for both students and teachers as it involves the natural environment in which teaching and learning is already taking place. Even if the interventions included here were to be considered evidence-based practices with strong approval from teachers, it cannot be assumed that these practices are continuing consistently in the classroom after researchers leave (Cook & Cook, 2011). Teachers often are presented with new, innovative programs that involve additional trainings and practice. For these practices to be effective, teachers are recommended to receive continual professional development throughout a school year on a target area (Daniel & Lemons, 2018). The research-to-practice gap continues to be a hurdle in the field of education, and it has led to more recent research specifically on ways to both quicken this transition and continue addressing issues with sustainability (Grünke et al., 2021; Sexton & Rush, 2021). In order to advocate for more sustained professional development around a particular approach to interventions, there must first be enough research that supports the allocation of school district resources for such training. Specific math problem solving intervention packages, such as the *Solve It!* curriculum, have been studied through randomized controlled trials that support this (Montague et al., 2011; Montague et al., 2014). Cook and Cook (2011) advocate that not all successful intervention strategies must be performed on such large scales though, particularly when focusing on interventions that are targeting students with disabilities in one-on-one or small group settings. Grünke et al. (2021) advocate for the inclusion of case study reports in education research – like those already commonplace in medical research – that would read more like lesson plan implementations of research-based practices. Although there is a time and place for researchers to serve as interventionists, it is imperative that more single-case design researchers implement a direct training model presented in a way that translates more easily to the existing classroom setting. This additional research using a direct training model will increase the support behind academic interventions for students with disabilities, including those in this study.

Intervention Types. The studies included a fair representation of each of the six intervention types, ranging from 3 to 5 studies implementing strategies from each category. In looking at general cognitive problem-solving strategy variations, Schaefer Whitby's (2013) study implemented a larger problem-solving curriculum called *Solve It!*. It was the only study included in this meta-analysis to use this cognitive strategy approach. Although it was only represented once in this analysis, the *Solve It!* intervention is a well-researched approach through a large-scale multi-year randomized controlled trial and other group-design studies (Krawec et al., 2013; Montague & Applegate, 1993; Montague et al., 2011; Montague et al., 2014). Full curriculum materials are available for purchase through Exceptional Innovations publishing. Preliminary single-case design studies implementing the intervention with students with disabilities occurred outside the parameters of this study (Montague, 1992). The schema-based intervention (SBI) approach, which was implemented in four of the included single-case design studies, has also been studied through large-scale randomized controlled trials (Jitendra et al., 2016; Jitendra et al., 2017).

Most of the intervention approaches were studied by different research teams and over a span of at least ten years. The main exception, the MSBI approach, included five studies (i.e., Root, Cox, Davis, & Hammons, 2020; Root Cox, Gilley, & Wade, 2020; Root, Cox, Saunders, & Gilley, 2020; Root et al., 2018; Root et al., 2017) that have all been published within the last five years. MSBI is designed for individuals with intellectual disabilities, a population that may not respond as well to less intensive intervention approaches that have been previously researched.

Across the studies in this review, technology was used in a variety of ways ranging from including digital supplemental material (i.e., a specific graphic organizer; Sheriff & Boon, 2014) and a computer-based program (Shin & Bryant, 2017). The use of video clips, video prompting, and video modeling as components of an intervention have been extensively researched in teaching individuals with significant disabilities academic, adaptive, and communication skills (Almalki, 2020; Dueker & Cannella-Malone, 2019; Knight et al., 2018; Cannella-Malone et al., 2017; Wu et al., 2016). As previous barriers to technology continue to be minimized (e.g., access, self-efficacy; Çoklar & Tatli, 2021), its use is likely to increase in future studies.

Additional limitations exist regarding the variations in problem types included and the researcher-made probes. Because each study determined a different method in creating a pool of sample real-world math questions to use on dependent measure probes, the conclusions drawn here would be stronger if a consistent global measure was used across all studies. This sort of global measure does not currently exist in math content areas, although researchers are actively proposing and studying potential tools to fill this void (VanDerHeyden et al., 2022).

Problem Types. The problem types in the reviewed studies consisted of general algebraic computations embedded within a word problem. Studies that involved other applications of problem solving, such as geometry, were not included. According to the Common

Core State Standards sixth graders should be taught real-world word problem practice involve using rates and ratios with whole-number measurements and writing and solving two-step variable equations and inequalities based on real-world contexts involving all four operations (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Seventh grade standards consist of teaching students real-world problem solving with rational numbers (e.g., adding fractions and decimals), and eighth grade standards consists of teaching abstract linear equations and real-world applications of volume of various 3D objects (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). It should be noted that the Common Core State Standards (CCSS) initiative is currently losing steam across the country, as many states are in the process of removing the standards and replacing them entirely with their own standards (Cramer et al., 2021).

Half of the studies in this review were targeting middle school level standards that were more closely aligned with sixth grade than eighth grade Common Core State Standards. Eleven studies (50.0%) focused solely on either addition/subtraction problem types or multiplicative/vary problem types. Even if these problem types would not be considered working on grade-level material, studies determined students' need for instruction in whichever problem type that was targeted in the study. Studies that included intellectual disabilities focused on teaching foundational math skills of addition and subtraction before focusing on multiple problem types or multiple operation choices.

Dependent Measures. The dependent variable detailed mathematic problem-solving accuracy across all studies, but each study used different researcher-created items and resources. It is a noted limitation within meta-analyses of intervention strategies that exact materials used as a dependent measure are not the same (Peltier et al., 2018). Unlike widely accepted general outcome measures for reading, there has not been a widely used general outcome measure for mathematics (VanDerHeyden et al., 2022).

Finally, the effectiveness of the results would be most applicable if generalization and maintenance data supported the success that each intervention demonstrated during intervention phases. The evidence provided within studies related to these measures is severely limited. Studies that did include a generalization measure designed them widely differently from one another. Similar variability existed among timing and amount of maintenance data gathered, making potential conclusions drawn on this evidence impossible without greater consistency across a larger sample size.

Self-Regulated Learning Components. Self-regulated learning (SRL) outcomes were not included in most of the studies that were reviewed, and therefore limited the analysis of the overall effects of SRL components on math problem-solving performance. The components of SRL that were measurable to some extent based on study designs were the use of cognitive strategies (e.g., paraphrasing, visualizing out loud, hypothesizing, computing, and checking of work) and meta-cognitive strategies (e.g., self-monitoring, goal setting, and self-evaluation). Unfortunately, less than half the studies did measured how well the participants were engaging in self-regulation. This is a limitation in the literature as there is much information that could be gleaned from the intersection of how well students with disabilities perform on self-regulation measures and their overall performance on academic measures (Mason & Reid, 2018).

Quality Indicators

All studies met inclusion criteria using the weighted method of calculating quality agreement from Royer et al. (2017). One third of studies met inclusion criteria using absolute coding. Even though including studies with the weighted method may introduce greater risk of variation, more quality studies that demonstrate significant results were able to be included using the weighted method. The lack of including an adequate explanation or description of how a component of the study was conducted does not automatically mean that it did not happen when implemented. Clear reporting is necessary for important replication work of intervention studies, but to exclude a study for potentially not reporting a single indicator component or two could restrict the analysis' overall conclusions. The ideas of successful intervention practices presented is much more robust when not implementing too rigorous of quality standards on this sample

In looking at the IRR data for the quality indicators, both overall IRR by component and overall IRR by article measures were in acceptable ranges above 90%. Five of the six studies that had less than 90% interrater agreement per article were published before the standards were in 2014 and the sixth article was published in 2015. When removing all studies that were published prior to 2014, the mean IRR per article increased to 96.5%. Although being published after the standards was not a requirement for inclusion, the standards have had an impact on the clarity of reporting results in single-case design research.

Finally, in looking at the mean IRR by component, the lowest agreement occurred with components that analyzed if a study's inclusion of implementation fidelity data related to dosage or exposure was adequate, if adequate fidelity data was included for each condition, if the design adequately controlled for threats to internal validity, and if frequency of measures were appropriate (as measured by a minimum of three data points per phase). All the studies included checklists or adequate measures of implementation fidelity related to adherence to intervention steps. As outlined in the standards though, if checklists are included for adherence, studies do not need to include additional data on dosage beyond outlining dosage amounts (Lane et al., 2014). Studies were not consistent in what type of dosage information was shared. Some included how long each session was but did not include how many sessions were conducted overall. Others detailed how many times per week sessions were held, but not how long each session lasted. Some only included over how many months the study took to complete. This was a point of confusion during coding when thorough information was provided about adherence, but inconsistent information was provided about dosage. Additionally, not all studies clarified if the adherence data that was included was split across all phase conditions or not. Some studies included data for different parts of the intervention package but did not include any data for baseline sessions. Both issues with implementation fidelity are necessary for appropriate levels of replication in future studies.

The next quality indicator component where there was less than 90% agreement was related to if the design of studies adequately controlled for threats to internal validity. Although multiple-baseline design is a commonly accepted single-case design study that automatically and adequately controls for threats to internal validity, all the studies included were multiple-probe baseline designs. There are tangible positives to choosing a multiple-probe design over a straight-forward multiple-baseline design in certain situations (Gast et al., 2018). In the cases of disagreement related to this in the analysis, there were questions as to whether researchers had adequately probed participants directly prior to introducing them to the intervention or if all participants had overlapping baseline data prior to anyone starting the intervention phase.

The final component with less than 90% agreement also related to graph analysis. The quality indicator outlined if frequency of measures were appropriate (as measured by a minimum of three data points per phase). Some graphs included sub-phase lines to indicate sub-components within the overarching intervention phase that did not include three data points per subphase. There were adequate data points overall within the whole intervention phase and no analyses were drawn from the sub-phase data, only the overall intervention package. Although these discrepancies quickly resolved themselves through conversation among the lead researcher and the independent coder, there are disputes among researchers regarding situations where more data points may be appropriate and necessary to draw the accurate conclusions (Ledford et al., 2018). Although all studies included adequate number of data points, there were multiple graphs that had questionable trends when the decision to intervene were made. Increasing the number of baseline data points for all participants would have been helpful in those instances.

Meta-Analysis Effect Size

The overall Tau-*U* results demonstrated a large effect of 0.93. This effect size specifically speaks to the effectiveness of cognitive strategy instruction within interventions targeting math problem-solving accuracy. More research would need to be conducted to draw clearer conclusions about the specific role other metacognitive strategies present in intervention designs have within these general effect sizes.

Three studies had moderate effect sizes less than 0.80. One involved a variation of the CRA intervention type and systematically worked to blend the models students used within each of the three instructional steps to represent problems (Scheuermann et al., 2009). This study was conducted with a high number of participants, meaning this may be a more reliable effect of the intervention within a small group. The next study involved a general approach that measured the effect of reading the word problems out loud, providing a graphic organizer, and direct instruction working through a task analysis of steps to solve the problems (Browder et al., 2012). Studies since Browder et al.'s that involve diagrams and graphic organizers have included more specific instructional components on proper use of a graphic organizer and diagram. Those studies have led to stronger results in math problem-solving with students with disabilities (i.e., Buock & Long, 2020; Buock et al. 2021). The final study with a moderate effect size utilized computer assisted instruction through the *Fun Fractions* computer program (Shin & Bryant, 2017). The researchers did not intervene or provide supplemental instruction in this study, so the effect size measured the sole effectiveness of students learning math problem-solving skills by completing modules through the computer program independently. These results are meaingful when considering students with disabilities often benefit more with supplemental instruction or support in addition to the computer modules on their own.

Moderator Analysis

The moderator analysis was completed for disability category, interventionist, group setting, intervention type, and problem type. Regarding disability category, all category groupings continued to have a large effect size. Students with intellectual disabilities had the highest effect size (Tau-U = 0.962) and students with ADHD/Other Health Impairment or an auditory processing disorder had the lowest effect size (Tau-U = 0.814). Since most of the studies that involved students with intellectual disabilities involved altering other effective practices specifically for the barriers of this group, it is encouraging that the effect size is so strong.

There was not a large difference between effect sizes for interventions implemented by a researcher or a special education teacher. The other two categories were based on the results of a single study and do not clearly add to the overall interpretation in this category. The effect size for students that participated in interventions conducted in a one-on-one format was higher than the effect size for students that participated in intervention conducted in a small group setting. This is not particularly surprising as feedback provided in an individual setting is more likely to be immediate and direct for the individual working versus a slightly more generalized approach within a small group. Regardless, both effect sizes continued to be measured as large effects. Similarly, both intervention types and problem types did not reveal any significant differences among subgroups and all effect sizes were greater than 0.80. Because of the variability introduced within the meta-analysis by including intervention approaches that were not identical along with a focus on different math skills within real-world math word problems, these moderator effects support the overall effect size by demonstrating that even when separated into matching interventions or matching skill focus, the effect sizes of participants were still large across the board.

Risk of Bias

The risk of bias assessment included in this meta-analysis is based on a framework presented by Reichow et al. (2018). Unfortunately, many of the other tools available to assess risk of bias consistently and accurately within studies of a meta-analysis sample are not designed for single-case design studies. For example, the Cochrane Risk of Bias tool, RoB2, is specifically for randomized trials, the Risk of Bias in Non-Randomized Studies of Interventions (ROBINS-I) is designed to compare effects of two or more interventions that target medical and health outcomes, and the Risk of Bias due to Missing Evidence (ROB-ME) tool adds components to the previous two in order to adequately address missing data in large data sets (The Cochrane Collection, 2022). Reichow et al.'s (2018) proposed framework and protocol tool of the SCD RoB therefore fills a necessary gap, but this protocol is not freely available. Therefore, the results of this bias assessment are limited within the framework of using Reichow et al.'s (2018) descriptions of categories within the tool on a researcher created template instead. It would have been ideal to have used a tool that had more established validity, but the SCD RoB provided a means to assess risk in a way that aligned most closely with the Cochrane tools most often used in group design studies and that targeted risks of bias that are applicable to single-case design studies in general.

It should also be noted that the SCD RoB does include an extra category as a catch-all for any potential types of additional bias that were not incorporated into the eight domains reported. Other potential sources of bias in education research could be due to timing of school holidays and breaks causing longer pauses than intended between sessions or loss of certain data due to technology malfunctions (Reichow et al., 2018). These types of potential bias introduction were not tracked during this analysis as most did not share information related to these types of situations. This highlights a limitation when single-case design graphs have an x-axis that is simply labeled “sessions” without additional information or measurement of time between sessions.

Final limitations included in the study relate directly to design or reporting flaws that lowered ratings on quality indicator and risk of bias measures. For example, studies were not always clear about sequence generation (selection bias) and how participants were chosen to be introduced to intervention even if their baseline data was identical. This raises the question about if and how randomization within single-case design should be more wide-spread. Some studies also took different approaches when reporting how this decision was made as well. Some studies documented the participant with the least variability in baseline data entered the intervention phase first due to first reaching steady state responding, while others chose the student with the most variability to give them more time to benefit from the intervention. Additionally in selection bias, participant selection methods needed increased clarity. Most studies included a report of how participants were appropriate for the study by cross-checking their performance on the dependent measure as a part of the inclusion/exclusion criteria. Other studies measured this differently or inferred general performance based on more global measures without cross-checking until formal baseline probes were conducted.

Two areas of performance bias and one of detection bias also introduced limitations to the evidence within studies. The blinding of participants and personnel (performance bias) and blinding of outcome assessors (detection bias) both were unclear in many studies. With single-case design studies, oftentimes the researcher is implementing the intervention, conducting probe measures, and analyzing data for future study decisions. It was not common in the studies included within this analysis to have independent coders complete fidelity and intercoder agreement measures on 100% of sessions and probes. This meant that blinding of personnel and outcome assessors was not fully assured. This does not appear to be common practice among single-case design studies but could be countered through explicit decision standards pre-set prior to any data collection being measured. This could in fact be what researchers were doing, but the report of procedure is unclear about when the cutoffs were determined. The final limitation related to performance bias introduced involved studies that did not report a fidelity measure or data for baseline phases. For many studies, the baseline phase only consisted of giving a student a worksheet, any allowed materials, and not providing any additional support. A fidelity checklist for this would be short and may seem unnecessary, but it is a crucial component to the reliability of the data to show that that no support is provided during this phase and that students received all materials required.

Implications of Results for Future Practice and Research

Future research related to math problem-solving interventions with middle-school students with disabilities should incorporate specific measures of SRL components within the interventions and in addition to academic outcomes. Few studies have been conducted looking specifically at the impact of SRL components on academic accuracy (Mason & Reid, 2018). These SRL measures should be defined and prepared in ways that align with the most current recommendations. For example, effective think-aloud protocols should include direct instruction and practice in the procedure prior to completing the task with the academic content in mind for the study, as well as adequate interobserver agreement (Greene et al., 2018). Trace data sources need to be adequately designed prior to the study with consideration for the time, level of detail within the sample, and context in which each piece of data will be gathered (Bernacki, 2018). Other approaches for inclusion of SRL components could utilize case studies to provide more tangible examples of implementation of problem-solving interventions as well as study the effect of included SRL practices on aspects of academic performance in a more targeted manner (Butler & Cartier, 2018; Grünke et al., 2021). Future researchers may consider including an analysis of the self-monitoring data (task checklists), goal setting and self-evaluation (graphs of progress). It would be interesting to conduct longitudinal studies with these aspects across elementary, middle, and high school student samples to see if the application of self-regulation skills impacts math problem solving performance differently at different stages of children's development.

Researchers are exploring ways of developing mathematical assessments that measures the mastery of math computation and problem-solving skills that are sequentially organized in a way in which one skill builds on another (VanDerHeyden et al., 2022). There are additional math problem solving interventions that have been researched or suggested that were not present in the current analysis. Future studies should aim to apply new strategies to the population targeted here as middle school students with disabilities. Similarly, the results of this study demonstrate a greater sample of students with intellectual disabilities and learning disabilities having exposure to the interventions. Additional studies could focus on students with emotional disturbances or attention-related disabilities more specifically to confirm if these intervention practices are as beneficial for students with those learning profiles. Further replication research of single-case design studies will strengthen overall conclusions that can be drawn from this research. Within disability groups, it may be useful to determine which type of math problem solving intervention is most effective. In an era where students in Generation Z have greater levels of digital nativity than any generation before them as well (Çoklar & Tatli, 2021), incorporating useful aspects of technology into education interventions is a realm of research that will likely continue to expand dramatically in upcoming decades.

Finally, other measures of effect size for single-case design could be also utilized in future studies, such as a between-case standardized mean difference (BC-SMD) measure that would allow for greater comparability to effect sizes gathered from group-design studies (Maggin et al., 2017). This would be particularly useful when conducting a meta-analysis that included both group-designs and single-case designs.

Conclusion

The purpose of this study was to examine the effectiveness of math problem-solving interventions with self-regulated learning components on the math problem-solving performance of middle school students with disabilities. A meta-analysis was conducted of all studies meeting inclusion/exclusion criteria between 2000-2021. Twenty-two studies were found through a systematic search of five prevalent databases that were accessible in English and met the following criteria: included a mathematic problem-solving intervention as the independent variable, included math problem-solving performance on real-world word problems as a dependent measure, included students aged 11-14 and/or were reported as being 6^{th}-8^{th} grade students, included disability qualification information for participating students, included a component of SRL, implemented a single-case research design, and passed a weighted measure of study quality. Independent raters conducted interrater agreement (IRR) coding on all steps of the systematic search, rating of quality indicator components for each study, and content analysis including SRL components. All mean IRR measures were higher than 85% prior to discussions between coders for final agreement.

Study characteristics revealed that most interventions have been conducted in a private space, one-on-one, and with a researcher as the interventionist. Students with a variety of disabilities were included in the study. Disabilities represented included intellectual disabilities, learning disabilities, Autism, ADHD/Other Health Impairment, and an auditory processing disorder. Studies took place in a variety of settings including public and private, urban and rural, Midwestern USA and southeast USA. Interventions involved six general categories of approaches including schema-based instruction, modified schema-based instruction, concrete-representation-abstract method variations, diagram variations, general cognitive strategy instructions, and technology-based approaches. Interventions similarly focused on different skills within the math problem-solving intervention ranging from solely addition and subtraction word problem types to curriculum-based one-variable equations embedded within the word problem. All studies included use of a cognitive or metacognitive strategy that related to self-regulated learning. Many others included components of self-management, goal setting, and self-evaluation. Some studies included additional dependent variables measuring diagram use, strategy use, and discrimination of problem type. All but two studies included social validity measures, either using open-ended interviews or questionnaires for students and/or teachers.

Generalization and maintenance measures were used inconsistently and with great variation across the studies in the sample.

The results of this study reported an overall large effect size for math problem-solving interventions with self-regulated learning components on math problem-solving performance for middle school students with disabilities. These results were supported through a variety of sensitivity checks. The moderator analysis did not reveal any significant differences across disability category, interventionist, group setting, intervention type, or problem type. The risk of bias assessment determined that few domains of bias and few studies had high risks of bias. Five of the domains had any high risk of bias present. None of the domains were fully rated as a low risk of bias.

Despite noted limitations, the results of this study reveal important implications for researchers and educators. Future studies conducted in this area should incorporate a more targeted approach to measuring the mediating effects of SRL components on math problem-solving intervention effectiveness by using well-designed think-aloud protocols, trace data sources, or case studies. Additional intervention approaches should be attempted with this population that have not been tried previously and should also be implemented targeting students with specific disabilities, such as emotional disturbance. Ultimately, the use of a different effect size measure that is more easily interpreted and not as impacted by potential inflation, particularly when comparing to widely accepted effect sizes within group design studies is also needed when conducting single-case design meta-analysis research.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Almalki, S. (2020). Using video modeling and video prompting to teach conversational skills to students with Autism: A consideration of effectiveness, practicality, and acceptability. International Journal of Early Childhood Special Education, 12(2), 103–114. https://doi.org/10.9756/INT-JECSE/V12I2.201062

Alter, P. (2012). Helping students with emotional and behavioral disorders solve mathematics word problems. Preventing School Failure, 56(1), 55–64. https://doi.org/10.1080/1045988x.2011.565283

Alter, P., Brown, E. T., & Pyle, J. (2011). A strategy-based intervention to improve math word problem-solving skills of students with emotional and behavioral disorders. Education and Treatment of Children, 34(4), 535–550. https://doi.org/10.1353/etc.2011.0028

Axe, J. B., Phelan, S. H., & Irwin, C. L. (2019). Empirical evaluations of Skinner's analysis of problem solving. Analysis of Verbal Behavior, 35(1), 39–56. https://doi.org/10.1007/s40616-018-0103-4

Babakhani, N. (2011). The effect of teaching the cognitive and meta-cognitive strategies (self-instruction procedure) on verbal math problem-solving performance of primary school students with verbal problem-solving difficulties. Procedia Social and Behavioral Sciences, 15, 563–570. https://doi.org/10.1016/j.sbspro.2011.03.142

Barrera-Mora, F., & Reyes-Rodriguez, A. (2019). Fostering middle school students' number sense through contextualized tasks. International Electronic Journal of Elementary Education, 12(1), 75–85. https://doi.org/10.26822/iejee.2019155339

Berger, J. L. (2009). The influence of achievement goals on metacognitive processes in math problem solving. Journal of Cognitive Education and Psychology, 8(2), 165–181. https://doi.org/10.1891/1945-8959.8.2.165

Bernacki, M. L. (2018). Examining the cyclical, loosely sequenced, and contingent features of self-regulated learning: Trace data and their analysis. In D. H. Schunk & J. A. Greene (Eds.), Handbook of Self-Regulation of Learning and Performance (2nd Ed., pp. 370–387). Routledge/Taylor & Francis Group.

Biswas, G., Baker, R. S., & Paquette, L. (2018). Data mining methods for assessing self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), Handbook of Self-Regulation of Learning and Performance (2nd Ed., pp. 388–404). Routledge/Taylor & Francis Group.

*Bouck, E. C., & Long, H. (2020). Teaching students with intellectual and developmental disabilities to calculate cost after discounts via schematic diagrams. Research in Developmental Disabilities, 102, 1–14. http://doi.org/10.1016/j.ridd.2020.103656

*Bouck, E. C., Long, H., & Park, J. (2021). Using schemas to support life skills mathematics for students with developmental disabilities. Education and Training in Autism and Developmental Disabilities, 56(1), 27–40.

*Bouck, E., Park, J., & Nickell, B. (2017). Using the concrete-representational-abstract approach to support students with intellectual disability to solve change-making problems. Research in Developmental Disabilities, 60, 24–36. http://doi.org/10.1016/j.ridd.2016.11.006

Brawand, A. (2013). Proportional reasoning word problem performance for middle school students with high-incidence disabilities (HID). [Doctoral dissertation, George Mason

University]. ProQuest Dissertations Publishing.

*Brawand, A., King-Sears, M. E., Evmenova, A. S., & Reagan, K. (2020). Proportional reasoning word problem perfromance for middle school students with high-incidence disabilities (HID). Learning Disability Quarterly, 43(3), 140–153. http://doi.org/10.1177/0731948719837920

Brossart, D. F., Laird, V. C., & Armstrong, T. W. (2018). Interpreting Kendall's Tau and Tau-U for single-case experimental designs. Cogent Psychology, 5(1), 1–26. https://doi.org/10.1080/23311908.2018.1518687

*Browder, D. M., Jimenez, B. A., & Trela, K. (2012). Grade-aligned math instruction for secondary students with moderate intellectual disability. Education and Training in Autism and Developmental Disabilities, 47(3), 373–388. https://www.jstor.org/stable/23879972

*Browder, D. M., Spooner, F., Lo, Y. Y., Saunders, A. F., Root, J. R., Ley Davis, L., & Brosh, C. R. (2018). Teaching students with moderate intellectual disability to solve word problems. The Journal of Special Education, 51, 222–235. https://doi.org/10.1177/0022466917721236

Bundock, K., Hawken, L. S., Kiuhara, S. A., O'Keeffe, B. V., O'Neill, R. E., & Cummings, M. B. (2021). Teaching rate of change and problem solving to high school students with high incidence disabilities at tier 3. *Learning Disability Quarterly, 44*(1), 35–49. https://doi.org/10.1177/0731948719887341

Butler, D. L., & Cartier, S. C. (2018). Advancing research and practice about self-regulated learning: The promise of in-depth case study methodologies. In Handbook of self-regulation of learning and performance, 2nd ed. (pp. 352–369). Routledge/Taylor &

Francis Group.

Cannella-Malone, H. I., Chan, J. M., & Jimenez, E. D. (2017). Comparing self-directed video prompting to least-to-most prompting in post-secondary students with moderate intellectual disabilities. International Journal of Developmental Disabilities, 63(4), 211–220. https://doi.org/10.1080/20473869.2017.1301695

Choi, J., & Walters, A. (2018). Exploring the impact of small-group synchronous discourse sessions in online math learning. Online Learning Journal, 22(4), 47–64. https://doi.org/10.24059/olj.v22i4.1511

Cleary, T. J., & Callan, G. L. (2018). Assessing self-regulated learning using microanalytic methods. In D. H. Schunk & J. A. Greene (Eds.), Handbook of Self-Regulation of Learning and Performance (2nd Ed., pp. 338–351). Routledge/Taylor & Francis Group.

Çoklar, A. N., & Tatli, A. (2021). Examining the digital nativity levels of digital generations: From Generation X to Generation Z. Shanlax International Journal of Education, 9(4), 433–444. https://doi.org/10.34293/education.v9i4.4224

Common, E. A., Lane, K. L., & Royer, D. J. (2017). Research to inform practice: Group comparison and single-case research design quality indicator matrix using Council for Exceptional Children 2014 standards check for understanding (CFU). Unpublished tool. Retrieved from https://www.ci3t.org/practice

Cook, B. G., & Cook, S. C. (2011). Unraveling evidence-based practices in special education. The Journal of Special Education, 47(2), 71–82. https://doi.org/10.1177/0022466911420877

Cook, S. C., Collins, L. W., Morin, L. L., & Riccomini, P. J. (2020). Schema-based instruction for mathematical word problem solving: An evidence-based review for students with learning disabilities. Learning Disability Quarterly, 43(2), 75–87. https://doi.org/10.1177/0731948718823080

Council for Exceptional Children. (2014). Council for Exceptional Children standards for evidence-based practices in special education. https://cecpioneers.exceptionalchildren.org/sites/default/files/2021-04/EBP_FINAL.pdf

Cox, S. K., & Root, J. R. (2020). Modified schema-based instruction to develop flexible mathematics problem-solving strategies for students with autism spectrum disorder. *Remedial and Special Education*, *41*(3), 139–151. https://doi.org/10.1177/0741932518792660

Cozza, B., & Oreshkina, M. (2013). Cross-cultural study of cognitive and metacognitive processes during math problem solving. School Science and Mathematics, 113(6), 275–284. https://doi.org/10.1111/ssm.12027

Cramer, E. D., Gallo, R. F., Salum, C., Munoz, L. R., Lafont, C. P., & Baptiste, B. B. (2021). Challenges of urban educators in implementing state standards for students with disabilities. Education and Urban Society, 53(9), 1001–1018. https://doi.org/10.1177/00131245211004564

Cuenca-Carlino, Y., Freeman-Green, S., Stephenson, G. W., & Hauth, C. (2016). Self-regulated strategy development instruction for teaching multi-step equations to middle school students struggling in math. The Journal of Special Education, 50(2), 75–85. https://doi.org/10.1177/0022466915622021

Daniel, J., & Lemons, C. (2018). Teacher perspectives on intervention sustainability: Implications for school leadership. School Leadership & Management, 38(5), 518–538. https://doi.org/10.1080/13632434.2018.1439465

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. Educational Psychology Review, 28(3), 425–474. https://doi.org/10.1007/s10648-015-9320-8

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. Metacognition Learning, 3, 231–264. https://doi.org/10.1007/s11409-008-9029-x

Dueker, S. A., & Cannella-Malone, H. I. (2019). Teaching addition to students with moderate disabilities using video prompting. Journal of Special Education Apprenticeship, 8(2), 1–20.

Efklides, A., Schwartz, B. L., & Brown, V. (2018). Motivation and affecti in self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), Handbook of Self-Regulation of Learning and Performance (2nd Ed., pp. 64–82). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315697048-2

Erez, G., & Peled, I. (2001). Cognition and metacognition: Evidence of higher thinking. Education and Training in Mental Retardation and Developmental Disabilities, 36(1), 83–93. https://www.jstor.org/stable/24481616

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology, 63*(3), 665-694. https://doi.org/10.1348/000711010X502733

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2014). How to design and evaluate research in education (9th ed.). McGraw-Hill Education.

*Freeman-Green, S. M., O'Brien, C., Wood, C. L., & Hitt, S. B. (2015). Effects of the SOLVE strategy on the mathematical problem solving skills of secondary students with learning disabilities. Learning Disabilities Research and Practice, 30(2), 76–90.

http://doi.org/10.1111/ldrp.12054

Fuchs, L. S., Gilbert, J. K., Powell, S. R., Cirino, P. T., Fuchs, D., Hamlett, C. L., … Tolar, T. D. (2016). The role of cognitive processes, foundational math skill, and calculation accuracy and fluency in word-problem solving versus prealgebraic knowledge. Developmental Psychology, 52(12), 2085–2098. https://doi.org/10.1037/dev0000227

Fuchs, L. S., Powell, S. R., Cirino, P. T., Schumacher, R. F., Marrin, S., Hamlett, C. L., … Changas, P. C. (2014). Does calculation or word-problem instruction provide a stronger route to prealgebraic knowledge? Journal of Educational Psychology, 106(4), 990–1006. https://doi.org/10.1037/a0036793.supp

Fuentes, E. B. (2019). Self-regulated learning, math achievement and self-efficacy of high school students with learning disabilities. [Doctoral dissertation, Fordham University]. ProQuest Dissertations Publishing.

Gast, D. L, Lloyd, B. P., & Ledford, J. (2018). Multiple baseline and multiple probe designs. In J. R. Ledford & D. L. Gast (Eds.), Single case research methodology (pp. 239-281). Routledge.

Greene, J. A., Deekens, V. M., Copeland, D. Z., & Yu, S. (2018). Capturing and modeling self-regulated learning using think-aloud protocols. In D. H. Schunk & J. A. Greene (Eds.), Handbook of Self-Regulation of Learning and Performance (2nd Ed., pp. 323–337). Routledge/Taylor & Francis Group.

Grünke, M., Hord, C., & Macdonald, S. J. (2021). A rationale for the use of case reports in special education: The significance of detailed descriptions of assessment and intervention scenarios for bridging the research-to-practice gap. Insights into Learning Disabilities, 18(1), 91–101. http://sure.sunderland.ac.uk/id/eprint/13514

Hawkins, S. M., & Heflin, L. J. (2011). Increasing secondary teachers' behavior-specific praise using a video self-modeling and visual performance feedback intervention. Journal of Positive Behavior Interventions, 13(2), 97–108. https://doi.org/10.1177/1098300709358110

Hord, C., & Marita, S. (2014). Students with learning disabilities tackle multistep problems. Mathematics Teaching in the Middle School, 19(9), 548–555. http://doi.org/10.5951/mathteacmiddscho.19.9.0548

Hoyle, R. H., & Dent, A. L. (2018). Developmental trajectories of skills and abilities relevant for self-regulation of learning and performance. In D. H. Schunk & J. A. Greene (Eds.), Handbook of Self-Regulation of Learning and Performance (2nd Ed., pp. 49–63). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315697048-2

Irwin, V., Zhang, J., Wang, X., Hein, S., Wang, K., Roberts, A., York, C., Barmer, A., Bullock Mann, F., Dilig, R., & Parker, S. (2021). Report on the condition of education 2021 (NCES 2021-144). U.S. Department of Education. National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.

*Jitendra, A. K., DiPipi, C. M., & Perron-Jones, N. (2002). An exploratory study of schema-based word-problem-solving instruction for middle school students with learning disabilities: An emphasis on conceptual and procedural understanding. Journal of Special Education, 36(1), 23–38. https://doi.org/10.1177/00224669020360010301

Jitendra, A. K., Dupuis, D. N., Star, J. R., & Rodriguez, M. C. (2016). The effects of schema-based instruction on the proportional thinking of students with mathematics difficulties with and without reading difficulties. Journal of Learning Disabilities, 49(4), 354–367. http://doi.org/10.1177/0022219414554228

Jitendra, A. K., George, M. P., Sood, S., & Price, K. (2010). Schema-based instruction: Facilitating mathematical word problem solving for students with emotional and behavioral disorders. Preventing School Failure, 54(3), 145–151. https://doi.org/10.1080/10459880903493104

Jitendra, A. K., Harwell, M. R., Dupuis, D. N., & Karl, S. R. (2017). A randomized trial of the effects of schema-based instruction on proportional problem-solving for students with mathematics problem-solving difficulties. Journal of Learning Disabilities, 50(3), 322–336. http://doi.org/10.1177/0022219416629646

Jitendra, A. K., & Star, J. R. (2011). Meeting the needs of students with learning disabilities in inclusive mathematics classrooms: The role of schema-based instruction on mathematical problem-solving. Theory Into Practice, 50(1), 12–19. http://doi.org/10.1080/00405841.2011.534912

Kazemi, F., Yektayar, M., & Bolban Abad, A. M. (2012). Investigation the impact of chess play on developing meta-cognitive ability and math problem-solving power of students at different levels of education. Procedia-Social and Behavioral Sciences, 32, 372–379. https://doi.org/10.1016/j.sbspro.2012.01.056

Kikas, E., Mädamürk, K., & Palu, A. (2020). What role do comprehension-oriented learning strategies have in solving math calculation and word problems at the end of middle school? British Journal of Educational Psychology, 90, 105–123. https://doi.org/10.1111/bjep.12308

Kingbeil, D. A., Van Norman, E. R., McLendon, K. E., Ross, S. G., & Begeny, J. C. (2019). Evaluating Tau-U with oral reading fluency data and the impact of measurement error. Behavior Modification, 43(3), 413–438. https://doi.org/10.1177/0145445518760174

Knight, V. F., Kuntz, E. M., & Brown, M. (2018). Paraprofessional-delivered video prompting to teach academics to students with severe disabilities in inclusive settings. Journal of Autism and Developmental Disorders, 48, 2203–2216. https://doi.org/10.1007/s10803-018-3476-2

Korinek, L., & deFur, S. H. (2016). Supporting student self-regulation to access the general education curriculum. Teaching Exceptional Children, 48(5), 232–242. https://doi.org/10.1177/0040059915626134

Krawec, J. L. (2014). Problem representation and mathematical problem solving of students of varying math ability. Journal of Learning Disabilities, 47(2), 103–115. https://doi.org/10.1177/0022219412436976

Krawec, J., & Huang, J. (2017). Modifying a research-based problem solving intervention to improve the problem-solving performance of fifth and sixth graders with and without learning disabilities. Journal of Learning Disabilities, 50(4), 468–480. https://doi.org/10.1177/0022219416645565

Krawec, J., Huang, J., Montague, M., Kressler, B., & De Alba, A. M. (2013). The effects of cognitive strategy instruction on knowledge of math problem-solving processes of middle school students with learning disabilities. Learning Disability Quarterly, 36(2), 80–92. https://doi.org/10.1177/0731948712463368

Lane, K. L., Common, E. A., Royer, D. J., & Muller, K. (2014). Group comparison and single-case research design quality indicator matrix using Council for Exceptional Children 2014 standards. Unpublished tool. Retrieved from http://www.ci3t.org/practice

Ledford, J., Lane, J. D., & Tate, R. (2018). Evaluating quality and rigor in single case research. In J. R. Ledford & D. L. Gast (Eds.), Single case research methodology (pp. 365–392).

Routledge.

Lee, J. B., & Cherney, L. R. (2018). Tau-U: A quantitative approach for analysis of single-case experimental data inaAphasia. American Journal of Speech-Language Pathology, 27(1S), 495–503. https://doi.org/10.1044/2017_AJSLP-16-0197

*Ley Davis, L. (2016). Effects of peer-mediated instruction on mathematical problem solving for students with moderate/severe intellectual disability. [Doctoral dissertation, The University of North Carolina at Charlotte]. ProQuest Dissertations Publishing.

Lo, Y., Correa, V. I., & Anderson, A. L. (2015). Culturally responsive social skill instruction for latino male students. Journal of Positive Behavior Interventions, 17(1), 15–27. https://doi.org/10.1177/1098300714532133

Lobo, M. A., Moeyaert, M., Baraldi Cunha, A., & Babik, I. (2017). Single-case design, analysis, and quality assessment for intervention research. Journal of Neurologic Physical Therapy: JNPT, 41(3), 187–197. https://doi.org/10.1097/NPT.0000000000000187

Long, A. C. J., Miller, F. G., & Upright, J. J. (2019). Classroom management for ethnic-racial minority students: A meta-analysis of single-case design studies. *School Psychology, 34*(1), 1–13. https://doi.org/10.1037/spq0000305

Ma, X., Xin, Y. P., Tzur, R., Si, L., Yang, X., Park, J. Y., … Ding, R. (2014). The effect of an intelligent tutor on math problem-solving of students with learning disabilities. In Proceedings of the Joint Meeting of PME 38 and PME-NA 36 (4), 145–152. https://files.eric.ed.gov/fulltext/ED599909.pdf

*Maccini, P., & Ruhl, K. L. (2000). Effects of a graduated instructional sequence on the algebraic subtraction of integers by secondary students with learning disabilities. Education and Treatment of Children, 23(4), 465–489.

Maggin, D. M., Lane, K. L., & Pustejovsky, J. E. (2017). Introduction to the special issue on single-case systematic reviews and meta-analyses. Remedial and Special Education, 38(6), 323–330. https://doi.org/10.1177/0741932517717043

Marshall, H., Valentic, G., & Rasmussen, S. (2019). Embedded tutoring to enhance dialogic feedback and improve student self-regulation. Learning Assistance Review, 24(2), 87–97.

Mason, L. H., & Reid, R. (2018). Self-regulation: Implications for individuals with special needs. In D. H. Schunk & J. A. Greene (Eds.), Handbook of Self-Regulation of Learning and Performance (2nd Ed., pp. 473–484). Routledge/Taylor & Francis Group.

Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. Annals of Dyslexia, 53, 218–253. http://doi.org/10.1007/s11881-003-0011-7

Mevarech, Z. R., Verschaffel, L., De Corte, E., Schunk, D. H., & Greene, J. A. (2018). Metacognitive pedagogies in mathematics classrooms: From kindergarten to college and beyond. In D. H. Schunk & J. A. Greene (Eds.), Handbook of self-regulation of learning and performance (2nd Ed, pp. 109–123). Routledge/Taylor & Francis Group.

Meyer, M. R., & Diopoulos, G. (2002). Anchored learning in context. Mathematics Teaching in the Middle School, 8(1), 16-21. https://www.jstor.org/stable/41181220

Mikolajewicz, N., & Komarova, S. V. (2019). Meta-analytic methodology for basic research: A practical guide. Frontiers in Physiology, 10(203), 1–20. https://doi.org/10.3389/fphys.2019.00203

Montague, M. (1992). The effects of cognitive and metacognitive strategy instruction on the mathematical problem solving of middle school students with learning disabilities. Journal of Learning Disabilities, 25(4), 230–248.

Montague, M. (2006). Math problem solving for middle school students with disabilities. [Report]. The Access Center: Improving Outcomes for All Students K-8. https://tinyurl.com/kxvxhew

Montague, M., & Applegate, B. (2000). Middle school students' perceptions, persistence, and performance in mathematical problem solving. Learning Disability Quarterly, 23(3), 215–227. https://doi.org/10.2307/1511165

Montague, M., Enders, C., & Dietz, S. (2011). Effects of cognitive strategy instruction on math problem solving of middle school students with learning disabilities. *Learning Disability Quarterly, 34*(4), 262–272. https://doi.org/10.1177/0731948711421762

Montague, M., Krawec, J., Enders, C., & Dietz, S. (2014). The effects of cognitive strategy instruction on math problem solving of middle-school students of varying ability. Journal of Educational Psychology, 106(2), 469–481. https://doi.org/10.1037/a0035176

Morin, L. L., Watson, S. M. R., Hester, P., & Raver, S. (2017). The use of a bar model drawing to teach word problem solving to students with mathematics difficulties. Learning Disability Quarterly, 40(2), 91–104. https://doi.org/10.1177/0731948717690116

Mulcahy, C. A., & Krezmien, M. P. (2009). Effects of a contextualized instructional package on the mathematics performance of secondary students with EBD. Behavioral Disorders, 34(3), 136–150. https://doi.org/10.1177/019874290903400303

*Na, K-E. (2009). Th effects of schema-based intervention on the mathematical word problem solving skills of middle school students with learning disabilities. [Doctoral dissertation, The University of Texas at Austin]. ProQuest Dissertations Publishing.

National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Retrieved from https://www.nctm.org/Standards-and-Positions/Principles-and-Standards/

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). Common core state standards for mathematics. https://www.corestandards.org/Math/

Ness, B. M., & Middleton M. J. (2012). A framework for implementing individualized self-regulated learning strategies in the classroom. Intervention in School and Clinic, 47(5), 267–275. http://doi.org/10.1177/1053451211430120

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., …Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. The British Medical Journal, 372(71), 1–9. https://doi.org/10.1136/bmj.n71

Peltier, C. J., Vannest, K. J., & Marbach, J. J. (2018). A meta-analysis of schema instruction implemented in single-case experimental designs. Journal of Special Education, 52(2), 89–100. https://doi.org/10.1177/0022466918763173

Popham, M., Counts, J., Ryan, J. B., & Katsiyannis, A. (2018). A systematic review of self-regulation strategies to improve academic outcomes of students with EBD. Journal of Research in Special Education Needs, 18(4), 239–253. https://doi.org/10.1111/1471-3802.12408

Powell, S. R., & Fuchs, L. S. (2018). Effective word-problem instruction: Using schemas to facilitate mathematical reasoning. Teaching Exceptional Children, 51(1), 31-42.

Ramirez, G., Chang, H., Maloney, E. A., Levine, S. C., & Beilock, S. L. (2016). On the relationship between math anxiety and math achievement in early elementary school: The role of problem solving strategies. Journal of Experimental Child Psychology, 141, 83–100. https://doi.org/10.1016/j.jecp.2015.07.014

Reddy, L. A., Cleary, T. J., Alperin, A., & Verdesco, A. (2018). A critical review of self-regulated learning interventions for children with attention-deficit hyperactive disorder. *Psychology in the Schools, 55*(6), 609–628. https://doi.org/10.1002/pits.22142

Reichow, B., Barton, E. E., & Maggin, D. M. (2018). Development and applications of the single-case design risk of bias tool for evaluating single-case design research study reports. Research in Developmental Disabilities, 79, 53–64. https://doi.org/10.1016/j.ridd.2018.05.008

*Root, J. R., Cox, S. K., Davis, K., & Hammons, N. (2020). Contextualizing mathematical problem-solving instruction for secondary students with extensive support needs: A systematic replication. Research and Practice for Persons with Severe Disabilities, 45(5), 241–255. http://dx.doi.org/10.1177/1540796920949448

*Root, J. R., Cox, S. K., Gilley, D., & Wade, T. (2020). Using a virtual-representational-abstract integrated framework to teach multiplicative problem solving to middle school students with developmental disabilities. Journal of Autism and Developmental Disorders, 1–13. https://doi.org/10.1007/s10803-020-04674-2

*Root, J. R., Cox, S. K., Saunders, A., & Gilley, D. (2020). Applying the universal design for learning framework to mathematics instruction for learners with extensive support needs. Remedial and Special Education, 41(4), 194–206.

*Root, J. R., Henning, B., & Boccumini, E. (2018). Teaching students with autism and intellectual disability to solve algebraic word problems. Education and Training in

Autism and Developmental Disabilities, 53(3), 325–338.

*Root, J., Saunders, A., Spooner, F., & Brosh, C. (2017). Teaching personal finance mathematical problem solving to individuals with moderate intellectual disability. Career Development and Transition for Exceptional Individuals, 40(1), 5–14.

Rosenzweig, C., Krawec, J., & Montague, M. (2011). Metacognitive strategy use of eighth-grade students with and without learning disabilities during mathematical problem solving: A think-aloud analysis. Journal of Learning Disabilities, 44(6), 508–520. https://doi.org/10.1177/0022219410378445

Ross, S. W., & Sabey, C. V. (2015). Check-in check-out + social skills: Enhancing the effects of check-in check-out for students with social skills deficits. Remedial and Special Education, 36(4), 246–257. https://doi.org/10.1177/0741932514553125

Royer, D. J., Lane, K. L., & Common, E. A. (2017). Group comparison and single-case research design quality indicator matrix using Council for Exceptional Children 2014 standards: Standards overview and walk-through guide. Unpublished tool. Retrieved from http://www.ci3t.org/pratice

*Saunders, A. F., Spooner, F., & Davis, L. L. (2018). Using video prompting to teach mathematical problem solving of real-world video-simulation problems. Remedial & Special Education, 39(1), 53–64. http://doi.org/10.1177/0741932517717042

Scammacca, N. K., Roberts, G. J., Cho, E., Williams, K. J., Roberts, G., Vaughn, S. R., & Carroll, M. (2016). A century of progress: Reading interventions for students in grades 4-12, 1914-2014. Review of Educational Research, 86(3), 756–800. https://doi.org/ 10.3102/0034654316652942

*Schaefer Whitby, P. J. (2013). The Effects of Solve It! on the Mathematical Word Problem

Solving Ability of Adolescents With Autism Spectrum Disorders. Focus on Autism & Other Developmental Disabilities, 28(2), 78–88. https://doi.org/10.1177/1088357612468764

*Scheuermann, A. M., Deshler, D. D., & Shumaker, J. B. (2009). The effects of the explicit inquiry routine on the performance of students with learning disabilities on one-variable equations. Learning Disability Quarterly, 32(2), 103–120. http://doi.org/10.2307/27740360

Schmitz, E. A., Jansen, B. R. J., Wiers, R. W., & Salemink, E. (2019). Do implicitly measured math-anxiety associations play a role in math behavior? Journal of Experimental Child Psychology, 186, 171–188. https://doi.org/10.1016/j.jecp.2019.05.013

Schunk, D. H., & Greene, J. A. (2018). Historical, contemporary, and future perspectives on self-regulated learning and performance. In D. H. Schunk & J. A. Greene (Eds.), Handbook of Self-Regulation of Learning and Performance (2nd Ed., pp. 1–15). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315697048-2

Sexton, S., & Rush, D. D. (2021). Roadmaps for reflection: Implementation drivers to bridge the research-to-practice gap in early childhood intervention. International Journal of Early Childhood Special Education, 13(2), 15–22. https://doi.org/10.9756/INT-JECSE/V13I2.211034

*Sheriff, K. A., & Boon, R. T. (2014). Effects of computer-based graphic organizers to solve one-step word problems for middle school students with mild intellectual disability: a preliminary study. Research in Developmental Disabilities, 35(8), 1828–1837. http://doi.org/10.1016/j.ridd.2014.03.023

Shin, M., & Bryant, D. P. (2015). Fraction interventions for students struggling to learn mathematics: A research synthesis. Remedial and Special Education, 36(6), 374–387. https://doi.org/10.1177/0741932515572910

*Shin, M., & Bryant, D. P. (2017). Improving the fraction word problem solving of students with mathematics learning disabilities. Remedial and Special Education 38(2). 76–86. http://doi.org/10.1177/0741932516669052

Sterne, J. A,C,, Savović, J., Page, M.J., Elbers, R.G., Blencowe, N.S., Boutron, I., Cates, C.J., Cheng, H.-Y., Corbett, M.S., Eldridge, S.M., Hernán, M.A., Hopewell, S., Hróbjartsson, A., Junqueira, D.R., Jüni, P., Kirkham, J.J., Lasserson, T., Li, T., McAleenan, A., Reeves, B.C.,… Higgins, J.P.T. (2019). RoB 2: a revised tool for assessing risk of bias in randomised trials. The British Medical Journal, 366(14898), 1–8. https://doi.org/10.1136/bmj.l4898

Swanson, H. L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. Journal of Educational Psychology, 100(2), 343–379. https://doi.org/10.1037/0022-0663.100.2.343

Tan, P. (2016). Developing mathematical understanding and students with emotional and behavioural disorders: A review of the literature. Emotional and Behavioural Difficulties, 21(4), 361–376. https://doi.org/10.1080/13632752.2016.1201639

The Cochrane Collection. (2022). Cochrane methods: Bias Resources. https://methods.cochrane.org/bias/resources

The Science of Math. (n. d.) What is explicit instruction? https://www.thescienceofmath.com/what-is-explicit-instruction

Tolar, T. D., Fuchs, L. S., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2016). Cognitive profiles of mathematical problem solving learning disability for different definitions of disability.

Journal of Learning Disabilities, 49(3), 240–256. https://doi.org/10.1177/0022219414538520

Tornare, E., Czajkowski, N. O., & Pons, F. (2015). Children's emotions in math problem solving situations: Contributions of self-concept, metacognitive experiences, and performance. Learning and Instruction, 39, 88–96. https://doi.org/10.1016/j.learninstruc.2015.05.011

Trezise, K., & Reeve, R. A. (2018). Patterns of anxiety in algebraic problem solving: A three-step latent variable analysis. Learning and Individual Differences, 66, 78–91. https://doi.org/10.1016/j.lindif.2018.02.007

United States Census Bureau. (2021). QuickFacts: United States (V2021) [Online tool]. https://www.census.gov/quickfacts

Usher, E. L., & Schunk, D. H. (2018). Social cognitive theoretical perspective of self-regulation. In D. H. Schunk & J. A. Greene (Eds.), Handbook of Self-Regulation of Learning and Performance (2nd Ed., pp. 19–35). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315697048-2

Van de Walle, J. A. (2007). Elementary and middle school mathematics: Teaching developmentally (6th ed.). Pearson. http://hdl.handle.net/2027/mdp.39015076110017

VanDerHeyden, A. M., Burns, M. K., Peltier, C., & Codding, R. S. (2022). The science of math: The importance of mastery measures and the quest for a general outcome measure. Communique, 50(5), 1 and 28–30.

VanDerHeyden, A. M., Hughes, E. M., Codding, R. S. (2021). The science of math and class-wide math intervention. Communique, 50(4), 1 and 26–29.

*van Garderen, D. (2007). Teaching students with LD to use diagrams to solve mathematical word problems. Journal of Learning Disabilities, 40(6). 540–553.

http://doi.org/10.1177/00222194070400060501

van Garderen, D. (2008). Middle school special education teachers' instructional practices for solving mathematical word problems: An exploratory study. Teacher Education & Special Education, 31(2), 132–144. http://doi.org/10.1177/088840640803100206

van Garderen, D., & Scheuermann, A. M. (2015). Diagramming word problems: A strategic approach for instruction. Intervention in School and Clinic, 50(5), 282–290. http://doi.org/10.1177/1053451214560889

Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling & Development, 93*(4), 403–411. https://doi.org/10.1002/jcad.12038

Vannest, K.J., Parker, R.I., Gonen, O., & Adiguzel, T. (2016). Single Case Research: web based calculators for SCR analysis. (Version 2.0) [Web-based application]. College Station, TX: Texas A&M University. Available from http://www.singlecaseresearch.org

Villeneuve, E. F., Hajovsky, D. B., Mason, B. A., & Lewno, B. M. (2019). Cognitive ability and math computation developmental relations with math problem solving: An integrated, multigroup approach. School Psychology, 34(1), 96–108. https://doi.org/10.1037/spq0000267.supp

Wang, A. Y., Fuchs, L. S., & Fuchs, D. (2016). Cognitive and linguistic predictors of mathematical word problems with and without irrelevant information. Learning and Individual Differences, 52, 79–87. https://doi.org/10.1016/j.lindif.2016.10.015

Wang, Y., & Sperling, R. A. (2020). Characteristics of effective self-regulated learning interventions in mathematics classrooms: A systematic review. Frontiers in Education, 5(58), 1–17. https://doi.org/10.3389/feduc.2020.00058

Weinstein, C. E., Acee, T. W., & Jung, J. (2011). Self-regulation and learning strategies. New Directions for Teaching and Learning, 126, 45–53. https://doi.org/10.1002/tl

Wilkerson, T. (2022, February). Developing a deep understanding of mathematics. National Council of Teachers of Mathematics. https://www.nctm.org/News-and-Calendar/Messages-from-the-President/Archive/Trena-Wilkerson/Developing-a-Deep-Understanding-of-Mathematics/

Winne, P. H. (2018). Cognition and metacognition within self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), Handbook of Self-Regulation of Learning and Performance (2nd ed., pp. 36–48). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781315697048-3

Wolters, C. A., & Won, S. (2018). Validity and the use of self-report questionnaires to assess self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), Handbook of Self-Regulation of Learning and Performance (2nd Ed., pp. 307–322). Routledge/Taylor & Francis Group.

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). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/MPS_PG_043012.pdf

Wu, P-F., Cannella-Malone, H. I., Wheaton, J. E., & Tullis, C. A. (2016). Using video prompting with different fading procedures to teach daily living skills: A preliminary examination. Focus on Autism and Other Developmental Disabilities, 31(2), 129–139. https://doi.org/10.1177/1088357614533594

Xin, Y. P. (2008). Use of schema-based instruction in solving mathematics word problems: An emphasis on prealgebraic conceptualization of multiplicative relations. Journal of Research in Mathematics Education, 39(5), 526–551.

Zhang, D. Z., Stecker, P., Huckabee, S., & Miller, R. (2016). Strategic development for middle school students struggling with fractions: Assessment and intervention. *Journal of Learning Disabilities, 49*(5), 515–531. https://doi.org/10.1177/0022219414562281

Zheng, X., Flynn, L. J., & Swanson, H. L. (2012). Experimental intervention studies on word problem solving and math disabilities: A selective analysis of the literature. Learning Disability Quarterly, 36(2), 97–111. https://doi.org/10.1177/0731948712444277

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), Handbook of Self-Regulation (pp. 13–41).

Appendix A

Single Case Design Risk of Bias Tool Framework

Table 9

Single Case Design Risk of Bias Tool Framework by Reichow et al. (2018)

Area of bias and domain	Description
Selection bias	The systematic differences between baseline characteristics of the participants compared.
Sequence generation	The procedures used to allocate participants to intervention conditions or the order of the conditions to which participants are exposed.
Participant selection	The criteria and process used to include and select participants appropriate for the research.
Performance bias	The systematic differences between participants in the care provided or in the exposure to factors other than the intervention under investigation.
Blinding of participants and personnel	The methods used to ensure members of the research team remain unaware of when the intervention is implemented to whom.
Procedural fidelity	The quality of the description for each experimental condition and the reporting of evidence indicating sufficient adherence to the intervention under investigation. This includes the procedures used to train implementers.
Detection bias	The systematic differences between participants in how outcomes are determined.
Blinding of outcome assessors	The procedures used to ensure the individuals collecting outcome data are unaware of the study conditions and research purpose.
Selective outcome reporting	The completeness of the data reported for all participants who began the study including those who withdrew and for each of the dependent variables.
Dependent variable reliability	The procedures and reporting of agreement or reliability indices for the outcome variables.
Data sampling	The extent to which the amount data collected for the research was sufficient to determine the level and trend of the data patterns in each condition to support the determination of a functional relation.

Note. From Development and applications of the single-case design risk of bias tool for evaluating single-case design research study reports by B. Reichow, E. E. Barton and D. M. Maggin, 2018, Research in Developmental Disabilities, 79, p. 53–64 (http://www.doi.org/10.1016/j.ridd.2018.05.008). Copyright Elsevier Ltd. Reprinted with permission; permission conveyed through Copyright Clearance Center, Inc.