The Effects of Early Numeracy Interventions for Students in Preschool and Early Elementary: A Meta-Analysis

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Gena Nelson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. Kristen McMaster, Adviser

December 2017

Acknowledgements

I would like to thank my adviser, Dr. Kristen McMaster. Before Kristen was my adviser, she was my doctoral seminar instructor, and in that capacity she provided thoughtful and constructive feedback that made me a stronger writer and a more critical thinker. As my adviser she effortlessly displayed kindness and respect, two qualities that I strive for in my professional and personal relationships.

I first applied to graduate school at the University of Minnesota to pursue a degree in school psychology, and at that time I began working with Dr. Ted Christ. While working alongside him I developed an excitement for research. Although I did not continue pursuing a degree in school psychology, Ted did not give up on me as a student and continued to advise me; for that, I am grateful.

I would also like to thank Dr. Michael Harwell for sharing his expertise and allowing me to learn from him. I am also extremely grateful to Dr. Michèle Mazzocco, who modeled professionalism and set high standards for research. I gained invaluable skills and knowledge while working in her research lab. I hope to transfer these skills to my own research and mentoring of future researchers.

Finally, I would like to acknowledge my writing group members, Steph Brunner, Jessie Kember, and Chloe Webb. Friendly competition, accountability, and monthly social evenings made it easier to see the light at the end of the tunnel.

Dedication

I dedicate my dissertation to my parents and my grandmother. I consider myself incredibly lucky to have parents who have always supported me as I have pursued my dreams. From the time I was a child, they set high expectations and taught me to value hard work. I am grateful to my dad who expects me to "do things right the first time," but also to my mom, who is always there to remind me that it is okay to make mistakes. I also dedicate my dissertation to my grandma Amy. She was tenacious and was never afraid to vocalize her opinion. I attribute my strong-willed nature and my "never give up" attitude to her; two qualities that I know got me where I am today.

Abstract

The purpose of this meta-analysis was to examine the effectiveness of early numeracy interventions for young students, including students with disabilities or those at-risk for mathematics difficulty (MD). This study evaluated preschool, kindergarten, and first-grade interventions on early numeracy content, instructional features, and methodological components that improved students' mathematics achievement. A total of 33 studies met inclusion criteria for this meta-analysis, with 51 treatment groups. Excluding outliers, the average weighted effect size for numeracy interventions across 49 treatment groups was moderate (g = 0.63), and the 95% confidence interval did not include zero [0.50, 0.73]. Results indicated that early numeracy interventions that included preschool and kindergarten students produced larger treatment effects than interventions with first-grade participants; in addition, treatment effects were slightly higher on average for students identified as at-risk for MD according to low socio-economic status and performance greater than the 25th percentile on a mathematics screener, compared to students who were identified as typically achieving or at-risk for MD according to performance below the 25th percentile. The results of the final meta-regression model for the total sample of studies indicated that the following predictors accounted for the most between-studies variance: concrete-representational-abstract instructional framework, intervention duration, risk status of participants, and the inclusion of counting with one-to-one correspondence in the intervention content (Pseudo R^2 = 75%). Directions for future research on conducting interventions are provided, and implications for educators implementing early numeracy interventions are discussed.

Table of Contents

List of Tables

List of Figures

Chapter I

Introduction

The purpose of this chapter is to provide an introduction to the current state of mathematics instruction and intervention for young children that provides the rationale for this meta-analysis. The research questions, a brief overview of the methodology, and the organization of this study are also reviewed. Finally, this chapter also provides definitions of key terms.

Young children enter school with diverse mathematical experiences and opportunities to learn mathematics. In fact, researchers have documented mathematics performance gaps among children at kindergarten entry regarding various prekindergarten factors. For example, children from low socio-economic status (SES) backgrounds have exhibited significantly lower levels of mathematical understanding in preschool and compared to peers from higher SES backgrounds (Starkey, Klein, & Wakeley, 2004), and children with access to center-based childcare in preschool have performed better in mathematics compared to children with access to only parental or family care (Magnuson, Meyers, Ruhm, & Waldfogel, 2004). Not only is it important to identify preschool and school-entry factors that may contribute to performance gaps, but it is also necessary to find ways to close these gaps. Poor mathematics achievement in later elementary grades is predicted by early numeracy skills measured in kindergarten (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004; Friso-van den Bos et al., 2015; Jordan, Kaplan, Ramineni, & Locuniak, 2009; Krajewski & Schneider, 2009), so early interventions that target numeracy skills are essential to closing performance gaps in mathematics.

One way that schools have attempted to close performance gaps is through the implementation of multi-tiered systems of support (MTSS). The MTSS framework considers mathematics instruction for all students, including students who receive special education services. When students have not mastered mathematics concepts, they may be considered for supplemental mathematics instruction through intervention in Tier 2 or Tier 3 settings (Powell & Fuchs, 2015). By providing extra mathematics instruction using evidence-based practices, schools hope to bolster the performance of students who are struggling in core instruction. Reviews of mathematics research support the use of interventions to increase mathematics understanding for all students, including students with disabilities and mathematics difficulty (MD). Across grade levels and mathematics content, these reviews highlight the use of multiple representations, explicit and systematic instruction, word problem solving strategy instruction, and computer-assisted instruction (Baker, Gersten, & Lee, 2002; Burns, Codding, Boice, & Lukito, 2010; Chodura, Kuhn, & Holling, 2015; Codding, Hilt-Panahon, Panahon, & Benson, 2009; Gersten, Chard, et al., 2009; Kroesbergen & Van Luit, 2003; Mononen, Aunio, Koponen, & Aro, 2015; Xin & Jitendra, 1999; Zhang & Xin, 2012).

Problem Statement

Despite the rise in research and refined instructional recommendations related to mathematics learning for students (e.g., National Council of Teachers of Mathematics [NCTM], 2006), the results of a collection of longitudinal studies of achievement reveal the fact that students who enter school with poor mathematics skills and understanding tend to have slower or parallel growth compared to their peers, often resulting in even larger gaps at the end of each school year (Jordan & Hanich, 2003; Jordan et al., 2009;

Judge & Watson, 2011; Nelson & Powell, in press; Vukovic, 2012). This achievement gap is especially concerning in mathematics because, as students progress through school, mathematics performance influences several long-term outcomes including enrollment in and completion of post-secondary school, preparedness for college coursework and the workforce, and employment outlook and hourly earnings in adulthood (Lee, 2012; Murnane, Willett, Braatz, & Duhaldeborde, 2001; Parsons & Bynner, 1997; Spielhagen, 2006). Given that children enter school with varying experiences in mathematics and that children who have low mathematics understanding at school entry continue to perform below peers throughout school, it makes sense that schools would choose to intervene early in mathematics. In order to have the greatest positive impact on mathematics achievement for students who require supplemental instruction or intervention, it is critical that researchers and practitioners understand features of effective interventions across each of the early numeracy domains and that effectiveness of early numeracy interventions may differ according to variables such as students' age or specific numeracy content.

Rationale and Significance

In recent years, early intervention and prevention of mathematics difficulties has become a primary concern for researchers and policy-makers, in large part due to consistently low achievement in mathematics. For example, over the last few decades, several reports providing recommendations for mathematics learning targets and standards have been released (NCTM's *An Agenda for Action* and *Focal Points*; *Mathematics Learning Early Childhood: Paths Toward Excellence and Equity* report; Common Core State Standards [CCSS]). Recently, there has also been an upward trend in research publications related to early numeracy (Methe et al., 2011); including a surge in empirical research designed to investigate the effectiveness of mathematics interventions for children in early childhood through first grade. Even with the uptake in intervention research, achievement gaps have remained in mathematics and longitudinal research indicates that students who lag behind peers at kindergarten entry continue to perform poorly through elementary school (Jordan et al., 2009; Judge & Watson, 2011). A comprehensive examination of the current literature would provide a clear sense of what is currently known about the effects of early numeracy interventions. Whereas conducting another intervention study would add to the literature base on early numeracy interventions, a meta-analysis will reveal the overall mean effect of numeracy interventions, and potentially identify key instructional components of interventions that influence the average treatment effect. The results of a meta-analysis on early numeracy interventions can provide clear direction for next steps in research and implications for designing numeracy interventions for young children.

Research Questions

The purpose of this meta-analysis was to review intervention research on early numeracy instruction for preschool, kindergarten, and first-grade students. Specifically, the meta-analysis answered the following research questions:

1. What is the overall mean effect, and how variable are those effects, of early numeracy interventions on mathematics outcomes for students in preschool, kindergarten, and first grade on proximal outcome measures (i.e., measures closely aligned to the intervention)? On distal outcome measures (i.e., general measures that are not specifically tied to the intervention)?

2. Which early numeracy domain is most investigated, and which domain produces the largest effect size?
3. What are the differential treatment effects of early numeracy interventions on mathematics outcomes across study characteristics (e.g., quality of the study), participant characteristics (e.g., typically achieving versus mathematics difficulty), and intervention characteristics (e.g., treatment length, intervention agent)?
4. Which variables (e.g., age, instructional format, intensity) account for the most between-studies variance for the total sample? For the three domains separately?

Overview of Methodology

Performing a meta-analysis allows researchers and practitioners to aggregate results from primary research to answer research questions regarding the effectiveness of an intervention or treatment of interest that represents a topic of interest (Borenstein, Hedges, Higgins, & Rothstein, 2009; Cooper, Hedges, & Valentine, 2009). After identifying the topic of interest (in this case, early numeracy interventions), the first step to conducting a meta-analysis is to retrieve relevant studies that meet inclusion criteria through a comprehensive literature search. In a meta-analysis, the features of each study or intervention represent the independent variables, while the effect size (treatment effectiveness) represents the outcome variable; thus, studies that meet inclusion criteria must be extensively coded for components that represent independent variables (e.g., participant demographics, intervention features, instructional features, effect sizes). After coding the studies, descriptive and meta-analytic data techniques are used. Descriptive analyses include summarizing information about the studies and treatment groups in an aggregated form, such as the total number of participants across studies and average year of publication. Meta-analytic techniques allow researchers to determine the overall weighted mean effect size and confidence interval, heterogeneity of the distribution of the effect sizes, and differences between subgroups of studies or treatment groups. Finally, a meta-regression allows researchers to explore the potential sources of variation in treatment effects. This series of analyses is conducted with the goal of synthesizing the research base on a topic such as early numeracy, and providing recommendations for practice and future research directions.

Definition of Key Terms

The purpose of this section is to define key terminology that I will use throughout this meta-analysis. First, I define key terms that are related to the mathematics area of numeracy. I will use the same three-structure domain of early numeracy employed by committees such as the National Research Council (NRC) and other researchers to describe preschool and early elementary numeracy interventions. I also define key terms related to instructional features. Then, I define key terms related to meta-analyses that are used throughout the Method and Results sections.

Numeracy terms. The three early numeracy domains, as well as skills within the domains are described here. Note that this is not a complete list of possible skills within each numeracy domain. Skills that were most frequently represented in the literature were selected for inclusion in this section.

Number. Students typically display understanding of skills in the Number domain by exhibiting informal mathematics knowledge. Informal mathematics skills are learned through play and daily interactions with the environment, and they are nontraditional or self-invented strategies that children use to navigate mathematical situations (Purpura, Baroody, & Lonigan, 2013). For the purposes of this meta-analysis, the following skills are categorized as Number domain skills:

Cardinality. Students have an understanding of the principle of cardinality when they understand that the last number in a count sequence represents the total quantity (Wynn, 1990). For example, if students are asked to count a set of four dots, and an examiner then asks, "How many dots?" students who have an understanding of cardinality will likely immediately respond, "four" while students who do not have an understanding of cardinality will likely need to recount the dots before responding to the examiner.

Counting error identification. This skill refers to the ability to identify errors in counting with and without objects. For example, students possess this skill if they are able to accurately identify double counting of objects, or missed number words when reciting the number sequence (e.g., "one, two, four, five, six").

Counting with one-to-one correspondence. This type of counting refers to matching each object in a set with a number word it is associated with while counting a set of objects. In other words, a child would count a set of objects without skipping objects, or double counting objects, and each object receives only one number-word label.

Numeral identification. Numeral identification refers to the ability to correctly match the number word or number name with the number symbol.

Subitizing. Subitizing refers to the ability to instantly recognize (i.e., subitize) quantities of up to three or four (Mandler, & Shebo, 1982).

Verbal counting. Verbal counting refers knowing the correct sequence of number words. This can be illustrated in different variations as well, such as counting forward from "one," counting forward from a number other than "one," and counting backward.

Relations. Difficulties in the development of Relations domain skills, such as number line representation, have implications for mathematics understanding that continue beyond the elementary grades (Geary, Hoard, Nugent, & Byrd-Craven, 2008). The following skills are categorized as Relations skills:

Magnitude comparison. Magnitude comparison refers to the ability to discern between the most (or "more") and least (or "less") value of two or more numerals.

Matching quantities, numbers words, and numerals. Match quantity refers to the action of matching equivalent numbers represented as either a set or quantity (e.g., dots), numeral, or number word.

Missing number. This skill refers to the ability to identify the missing number that comes before or after a number, or between two numbers when provided a segment of the number line.

Number line estimation. Number line estimation refers to the ability to estimate the placement of a number along a number line when considering the beginning and end point of the number line that is provided.

Number line sequencing. Number line sequencing refers to the ability to correctly sort numbers (such as numerals on cards) in a specified order, for example, from least to greatest.

Ordinal numbers. Skill with ordinal numbers refers to the ability to state an objects' relative placement. For example, this skill can be used when describing a person's place in line as the "first" person in line, or the "second" person in line.

Set comparison. Set comparison (similar to magnitude comparison) refers to the ability to discern between the most (or "more") and least (or "less") when considering two or more groups of objects.

Operations. Skills in the Operations domain such as putting numbers together and taking numbers apart are prerequisite skills for developing basic fact fluency. The following skills are categorized as Operations skills:

Composition and decomposition. Composition refers to the ability to put numbers together to make larger numbers, or combine more than one set of objects to create a larger set of objects. Conversely, decomposition refers to the ability to take apart numbers to make smaller numbers, or take apart a set of objects to create two or more smaller sets of objects. The processes of composition and decomposition happen without formal operational words (i.e., plus, minus) or symbols (i.e., +, –), but instead emphasizes part-whole relationships. For example, students may be presented with a set of objects, and after an examiner alters the set, they may ask the student, "How many objects were added?" or "How many objects were taken away?"

Equivalence. Skills with equivalence refer to creating an equivalent set (from a numeral or set of objects), and breaking apart an initial set of objects into two or more equivalent sets of objects.

Properties of addition and subtraction. This refers to conceptual understanding of mathematical properties that are necessary to be proficient with addition and subtraction (e.g., knowing that when you add one to a set the total quantity become greater).

Simple addition and subtraction with objects. Basic arithmetic refers to the ability to add and subtract, typically within 20 (i.e., sums and differences are between 0 and 20). This may also include number combination skills generally referred to as "plus one/minus one," or "two more/two less" or "doubles." Students perform these actions with the use of objects (such as blocks) to help track addition and subtraction processes.

Simple addition and subtraction without objects. This skill is similar to simple addition and subtraction described above, except that students perform these actions without the use of objects, but instead use numerals.

Instructional features. Common instructional components (details of how lessons were delivered, or the context in which lessons were delivered) that were coded in this meta-analysis are described here.

Behavior management plan. A behavior management plan refers to a system that is above and beyond typical reinforcement and feedback during the intervention period (e.g., group contingency plan).

Concrete representations. Concrete representations refer to physical manipulatives (e.g., blocks, counting chips) that students may use to solve mathematics problems.

Concrete-representational-abstract framework. The concrete-representational-abstract (CRA) framework refers to a model of instruction where students first utilize physical or concrete manipulatives to solve problems (e.g., using blocks to solve story problems); once students have mastered a skill using concrete manipulatives they may be ready to use pictorial representations (e.g., drawing pictures of blocks to solve story problem, using tallies). Finally, once students are proficient with a skill using pictorial representations, they may be ready to use abstract symbols to solve problems (e.g., creating a number sentence with numerals and symbols to solve a story problem; Witzel & Little, 2016; p.27).

Explicit and systematic instruction. Explicit and systematic instruction is marked with several features including teacher modeling, guided practice, multiple practice opportunities for learners, corrective feedback, intentional sequencing of lessons and topics (Witzel & Little, 2016; p. 25).

Pictorial representations. Pictorial representations refer to things such as drawings that students may use to solve mathematics problems.

Scripted lessons. Scripted lessons are intervention materials that allow for instruction to be delivered in a standardized format.

Meta-analytic terms. The following terms are used throughout the Method and Results and are related important components of meta-analyses.

Effect size. An effect size is a metric used in a meta-analysis to represent the magnitude of an intervention or treatment effect. Effect sizes allow treatment effects to be transformed to a common metric that enables researchers to include different outcome measures in the same meta-analysis or synthesis (Borenstein et al., 2009).

Fixed-effect and random-effects models. Researchers may choose between a fixed-effect model or random-effects model when conducting a meta-analysis. In a fixed-effect model, researchers assume that there is one true effect size that represents all studies in the analysis. Any observed differences in the effect size estimate are attributed to sampling error. In contrast, in a random-effects model, researchers assume that there may be different effect sizes that represent different studies and that the true effect may vary. Variation in effect sizes in a random-effects model is attributed to variables (e.g., differences between participants, intervention characteristics) in addition to sampling error (Borenstein et al. 2009).

Proportion of variance (I^2). The I^2 value represents the proportion of variability in effect sizes that was due to heterogeneity rather than sampling error (Borenstein et al., 2009; Cooper et al., 2009). The larger the proportion, the more variability that is attributed to *true* differences among treatment effects, compared to differences due to sampling error. An I^2 value near zero indicates that a large proportion of the variation in effect sizes is due to sampling error, while I^2 values greater than 75% are considered to represent high amounts of heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003).

Q test. The *Q* test is a test of significance and a measure of heterogeneity. *Q* tests are conducted to examine variation among the distribution of effect sizes and a statistically significant *Q* statistic represents a heterogeneous distribution of treatment effects. When an overall *Q* test is significant, further examination of potential causes of the variation is supported (Borenstein et al., 2009).

Organization of the Dissertation

This study reports the findings of a meta-analysis of primary studies that investigated the effectiveness of early numeracy interventions for young students. Chapter 2 discusses the theoretical foundations and historical trends guiding this study and a literature review of previous mathematics syntheses. Chapter 3 describes the methods used to conduct the literature review to identify relevant studies, code variables of interest, calculate effect sizes, and conduct data-analyses. Chapter 4 reports the descriptive and meta-analysis results. Chapter 5 frames the findings of this meta-analysis in the context of the current state of mathematics education and achievement; limitations of this study are discussed, suggestions for future research are made, and implications for educators of students who struggle in mathematics are reviewed.

Chapter II

Literature Review

The purpose of this chapter is to provide a review of the literature to support a meta-analysis on early numeracy interventions. In this chapter, I first review research findings regarding the development of numeracy skills in young children and the long-term importance of developing these skills. Then, I review mathematics education policy that has emerged as a result of consistently low achievement in mathematics by many students. Finally, I describe previous reviews that examined the effectiveness of mathematics interventions.

Development of Mathematics Knowledge

The acquisition of skills in the broad domain of mathematics starts long before children enter formal schooling. More specifically, most of the mathematics skills that develop early are categorized as numeracy skills. Early numeracy is broadly defined as the understanding of numbers and the relations between numbers (Malofeeva, Day, Savo, Young, & Ciancio, 2004). Research with infants indicates that people possess an innate sense of number, and that some capacity of understanding number is present in a pre-linguistic stage of development (Resnick, 1989; Dehaene, 2011). Infants are capable of remembering small sets of items (two to three objects) and discriminating between sets of items (Starkey & Cooper, 1980; Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005). When children become able to verbally express some understanding of mathematics concepts, they may recite part of the number sequence without any real purpose of counting or labeling objects (e.g., they may sing or chant number words with no pattern). This skill emerges around 2 to 3 years of age (Clements & Sarama, 2009). Children between the ages of 3 and 5 years accurately exhibit their ability to subitize a set of three objects. Subitizing is the ability to instantly and accurately recognize groups of objects (e.g., dot arrays) between one and three to four items. Subitizing is considered a nonserial—instant and automatic—process, which contrasts with a conscious counting procedure (Wynn, 1990).

As children experience more opportunities to interact with their environment, they begin to display understanding of other informal mathematics knowledge (i.e., nontraditional or self-invented strategies used to navigate mathematical situations) such as reciting the verbal count sequence, counting objects with one-to-one correspondence, and cardinality (Krajewski & Schneider, 2009; Purpura et al., 2013; Purpura & Lonigan, 2013). Knowledge of the verbal counting sequence typically develops throughout the preschool years, and by approximately 4 years of age most students are able to accurately count a small set of objects (e.g., blocks) and determine how many objects are in the set (Sarnecka & Carey, 2008; Wynn, 1990). As students' number skills progress, they establish cardinality, which confirms that they understand that the last number in a count sequence represents the total quantity of objects in a set (Wynn, 1990).

Informal mathematics knowledge is the foundation for the later development of formal mathematics skills; research also indicates that informal mathematics skills are strong predictors of later formal mathematics abilities (Jordan, Kaplan, Nabors Oláh, & Locuniak, 2006; Lembke & Foegen, 2009; Mazzocco & Thompson, 2005). Formal mathematics skills typically emerge once students enter school and receive mathematics instruction. When students have an understanding of formal mathematics knowledge, they are able to demonstrate skills such as writing numerals and numerical notation (e.g., using the “+” symbol). They also begin to display written and verbal fluency with basic facts. By kindergarten and first grade, students are expected to understand the relative position of cardinal numbers (NCTM, 2006). The development of skills such as understanding the relative position of a number or understanding how two or more numbers are connected are needed to develop more complex mathematics skills, such as problem solving. Although some children may develop skills such as simple addition and subtraction without formal instruction, these skills typically emerge with formal instruction in kindergarten and first grade. Proficiency in basic arithmetic facts is essential to solve word problems and complex computation problems (Fuchs et al., 2010).

Although numeracy skills develop prior to school entry for many children, researchers have found that children who entered kindergarten with deficits in early numeracy skills such as counting and magnitude comparison continued to perform poorly in mathematics throughout elementary school, and showed slower growth in mathematics compared to higher performing students (Bodovski & Farkas, 2007; Judge & Watson, 2011). Jordan et al. (2009) found that skills such as counting and number comparison measured in kindergarten were significant predictors of overall mathematics achievement at the end of third grade ($r = 0.63$; $p < .01$). Morgan, Farkas, and Wu (2009) observed that students who entered and exited kindergarten with mathematics achievement below the 10th percentile had a 70% chance of scoring below the 10th percentile again 5 years later. In a meta-analysis of six longitudinal studies, Duncan et al. (2007) determined that of the three school-entry variables (reading, mathematics, attention) to predict later reading and mathematics achievement, mathematics skills at school entry had the strongest association with later school achievement ($B = 0.33$; $p < .001$).

Collectively, results from developmental studies suggest that children who do not develop foundational numeracy skills before the end of kindergarten are likely to have difficulty with the attainment of higher-level mathematics skills in later grades (Nelson & Powell, in press). Thus, the development of numeracy skills in the early years of formal schooling is critical to children's success in mathematics (Geary et al., 2008; Magnuson et al., 2004; Starkey et al., 2004).

Policy Related to Mathematics

In light of longitudinal research that indicates the importance of foundational numeracy skills for later mathematics success and the consistently low achievement in mathematics, early intervention and prevention of mathematics difficulties has become a primary concern for researchers and policy-makers in recent years (Geary et al., 2008; Jordan et al., 2009; National Center for Educational Statistics, 2013a). Recently, there has also been an upward trend in research publications related to mathematics difficulties and intervention (Methe et al., 2011); however, the push for a greater focus on mathematics started several decades ago. In this section, I highlight policy reports and legislation that has emerged related to mathematics education and the acknowledgement of early numeracy skill development.

NCTM An Agenda for Action: 1980s. In response to a debate that spanned the 1960s and 1970s regarding mathematics education ("new math" versus "back-to-the basics"), the NCTM implemented the Priorities in School Mathematics (PRISM) project. The results of this project served as the basis for the NCTM report, *An Agenda for Action: Recommendations for School Mathematics for the 1980s* (NCTM, 1980). The report consisted of eight detailed recommendations for mathematics education.

Recommendations included emphasizing problem-solving skills, enhancing computation abilities beyond fluency, incorporating technology in teaching, measuring the effectiveness of teaching with more than conventional tests, employing stringent standards for effective mathematics instruction, requiring more mathematics learning for students, requiring mathematics teachers to display a high level of professionalism, and focusing society's attention on the importance and value of mathematical understanding. The report was widely distributed among teachers, parents, and policy-makers, and was featured in discussions of mathematics education during the 1980s (Center for the Study of Mathematics Curriculum [CSMC], 2004; NCTM, 1980).

Following *An Agenda for Action*, the National Commission on Excellence in Education (NCEE) released the *A Nation At Risk* (1983) report, and highlighted the fact that there was a substantial shortage of qualified mathematics teachers, mathematics curricula content was diluted, and states had lowered expectations regarding the number of mathematics courses students were required to take. The report raised concerns regarding the quality of mathematics instruction that students were receiving. After the releases of the *A Nation at Risk* and *An Agenda for Action*, NCTM released universal standards for learning in mathematics (CSMC, 2005; NCEE, 1983).

NCTM Standards: 1989. *An Agenda for Action* aimed to change the direction of mathematics education; subsequently, NCTM released the *Curriculum and Evaluation Standards for School Mathematics* (NCTM Commission on Standards for School Mathematics, 1989). The standards were developed by a commission, and the commission outlined several goals for students, including learning the value of mathematics, learning to communicate and reason mathematically, and becoming mathematical problem solvers. The commission also organized the standards by grade band. Specifically, the standards for the kindergarten to fourth (K–4) grade band emphasized a decrease in rote learning, with an increase in number sense skills and reasoning. The 1989 NCTM *Standards* also emphasized that K–4 mathematics curricula should include whole number concepts because students must first understand numbers if they want to make sense of numbers and mathematics around them in everyday contexts. The NCTM *Standards* also emphasized the importance of counting skills, composition and decomposition, understanding number magnitude, and developing an understanding of operations of whole numbers. These foundational whole number skills are needed for children to master more complex mathematics skills later, such as place value concepts, computation, and problem solving. The NCTM *Standards* provided a universal framework for expectations for mathematics learning and proficiency in the U.S. (CSMC, 2004; NCTM Commission on Standards for School Mathematics, 1989). Following the release of the 1989 *Standards*, NCTM released an updated version in 2000–*Principles and Standards for School Mathematics*. The updated 2000 NCTM *Standards* included recommendations for preschool learners as well as narrower age-bands to provide more details for specific skills to be taught at each grade (NCTM, 2000).

A longstanding critique of mathematics curricula in the U.S. has been that it is a "mile wide and an inch deep," and NCTM and the National Governors Association Center for Best Practices & Council of Chief State School Officers (NGA & CCSSO) released new sets of standards to address this critique in hopes of refining mathematics curricula (NGA & CCSSO, 2010; NCTM, 2006).

NCTM Focal Points: 2006. In 2006, NCTM published the *Curriculum Focal Points for Prekindergarten through Grade 8 Mathematics: A Quest for Coherence* (NCTM, 2006). The purpose of the *Focal Points* was to identify key mathematics topics at each level from prekindergarten to eighth grade and provide schools with a coherent framework for mathematics curriculum development. The report cited that at each grade level there were too many standards across curricula and state expectations, which created a lack of consensus and decreased emphasis on key ideas. For example, for fourth grade, the report identified a range of 26 to 89 different mathematics standards across 10 states.

As a result, for prekindergarten (PreK), kindergarten (K), and first grade (1), NCTM identified these three key areas for mathematics learning: Numbers and Operations (PreK, K, 1), Geometry (PreK, K, 1), Measurement (PreK, K), and Numbers and Operations and Algebra (1). Across the Numbers and Operations strand of the *Focal Points*, students are expected to develop an understanding of whole numbers, counting principles, comparison, order, composition and decomposition, and strategies for basic addition and subtraction. The aim of the *Focal Points* was to help schools align curriculum with a specific set of grade-level expectations in hopes of teachers being able to commit more time during the school year to those topics. Furthermore, because the *Focal Points* spanned prekindergarten through eighth grade, the use of the *Focal Points* by schools would allow for a connected and solid mathematics foundation (NCTM, 2006).

Common Core State Standards: 2010. Following the release of the NCTM *Focal Points,* the NGA & CCSSO (2010) released the *Common Core State Standards*

(CCSS) for mathematics and English and language arts. The CCSS was an educational initiative that aimed to establish consistent kindergarten to 12th grade standards for learning and ensure that students who graduated from high school were college or career ready. The CCSS identified key domains across grade levels and provided specific learning targets. In the kindergarten and first-grade mathematics standards, fewer topics were addressed with each topic receiving greater emphasis–consistent with the goals of the NCTM *Focal Points*. The kindergarten standards highlighted the importance of representing and comparing whole numbers, and describing shapes and spatial relations. In fact, the CCSS stated, "More learning time in Kindergarten should be devoted to number than to other topics" (p. 9). In first grade, the focus on whole numbers was extended to operations, understanding place value and number relationships, measurement, and geometry. The CCSS provided critical learning goals for what students should know and be able to do in kindergarten and first grade. Since the release of the CCSS, 42 states (as well as the District of Columbia and U.S. territories) have adopted the CCSS for mathematics (NGA & CCSSO, 2010).

Mathematics Learning Early Childhood: 2009. In addition to the NCTM and CCSS recommendations for students of all ages, in 2009 the Committee on Early Childhood Mathematics released the *Mathematics Learning Early Childhood: Paths Toward Excellence and Equity* report with recommendations specific for young children ages 3 to 6 years (NRC, 2009). The committee acknowledged that although mathematics education had recently been identified as a top priority in the educational policy agenda, many children continued to exhibit chronically low achievement in mathematics. The committee noted particular concerns for students from low socioeconomic status (SES) and the fact that achievement disparities exist even before students enter kindergarten. The committee made several recommendations, including that a national early childhood mathematics initiative be put in place, that all early childhood programs provide high-quality mathematics instruction as recommended by the report, and that education partnerships be formed between families and community programs to promote mathematics learning. In addition, the committee identified two critical areas that early childhood mathematics curricula should focus on, including: (a) numbers (including operations and relations), and (b) geometry, spatial relations, and measurement. The committee also acknowledged that more mathematics time should be devoted to numbers skills than to any other topic. The committee also published the report as a framework for principles and teaching-learning paths to be used in the classroom by teachers, as well as by curriculum developers for early childhood mathematics. The report includes recommendations that are developmentally appropriate learning targets and trajectories for mathematics.

Review of Previous Syntheses of Mathematics Interventions

Over the last few decades, several reports regarding recommendations for mathematics learning targets and standards have been released; concurrently, there has been a surge in the empirical research designed to investigate the effectiveness of mathematics interventions for children in early childhood through first grade. Even with the uptake in intervention research, achievement gaps have remained in mathematics and longitudinal research indicates that students who lag behind peers at kindergarten entry continue to perform poorly through elementary school. Determining effective instructional components within mathematics interventions may help refine recommendations for future interventions and instructional practices. A common practice to identify effective instructional practices for a particular domain is to conduct a synthesis of interventions. To date, researchers have published several meta-analyses, syntheses, and research reports related to mathematics.

Researchers have conducted syntheses and meta-analyses to examine the effectiveness of mathematics interventions for school-aged students, including typically developing students, students at-risk for mathematics difficulties (MD), and students with disabilities (Baker et al., 2002; Burns et al., 2010; Chodura et al., 2015; Codding et al., 2009; Gersten, Chard, et al., 2009; Kroesbergen & Van Luit, 2003; Mononen et al., 2015; Xin & Jitendra, 1999; Zhang & Xin, 2012). In this review of previous syntheses, I first discuss results from those studies that examined the effects of interventions for participants in kindergarten through 12th grade regarding broad mathematics, word problem solving, and computation. Then, I discuss results from syntheses that evaluated similar types of interventions that included only younger participants through 6th grade, followed by a description of reviews that exclusively focused on numeracy interventions. Then, I present findings from curriculum reviews specific to preschool mathematics. Finally, I discuss limitations of the previous syntheses. In the following sections, effect sizes are interpreted in the following manner: 0.20 = small, 0.50 = moderate, and 0.80 = large (Cohen, 1988).

Reviews for students through 12th grade. Several reviews have been conducted related to interventions for students in kindergarten through 12th grade for broad mathematics, word problem solving, and computation.

Broad mathematics. Baker et al. (2002) reviewed mathematics interventions for low-achieving students in kindergarten through 12th grade. The authors searched relevant literature from 1971 to 1999; 15 studies met inclusion criteria. Baker et al. (2002) used meta-analytic techniques to examine the variation in effect sizes on a study-by-study basis to evaluate the mean effect of interventions for specific categories of interest. The specific categories were: (a) providing data and feedback to teachers and/or students about performance, (b) peer-assisted learning strategies, (c) parent support of mathematics learning, (d) explicit or teacher-directed instruction, and (e) computer-assisted instruction. The authors also coded for the following variables: assignment to treatment condition, grade level, length of the intervention, fidelity, and percentage of students identified as low achieving or receiving special education services.

From the 15 included studies, 39 independent effect sizes were calculated and effects ranged from –0.59 to 1.49. The results indicated that interventions identified as having the following instructional characteristics had mean effect sizes significantly greater than zero: peer-assisted learning strategies (d = 0.62; 95% CI [0.42, 0.89]), explicit instruction (d = 0.65; 95% CI [0.40, 0.77]), and providing students with data about their performance (d = 0.71; 95% CI [0.27, 0.87]). In contrast, interventions that provided instructional recommendations to teachers, provided feedback to parents, or incorporated teacher facilitated instruction and practice yielded effects that were not significantly greater than zero.

Similarly, Gersten, Chard, et al. (2009) reviewed relevant literature between 1971 and 2007 that focused on mathematics interventions for students with documented learning disabilities. They included 42 interventions for participants in kindergarten through 12th grade. The authors examined several variables including mathematics domain and the nature of instruction. Specific mathematics domains included: operations, word problems, fractions, algebra, and general mathematics proficiency. Regarding the nature of instruction, authors coded information in four categories: (a) approaches to instruction, (b) providing feedback and/or data to teachers about student performance, (c) providing feedback and/or data to students about their performance, and (d) peer-assisted instructional strategies. The approaches to instruction category was broken down further and the following curricular design elements were also coded: explicit instruction, heuristics or general steps to solve a particular problem-type, student communication of mathematical reasoning process, visual representations, range and sequence or pattern of examples, and other.

Of all of the approaches to instruction characteristics, those with notably large mean random effects included explicit instruction ($g = 1.22$; 95% CI [0.78, 1.67]), use of heuristics ($g = 1.56$; 95% CI [0.65, 2.47]), and student communication about mathematical reasoning ($g = 1.04$; 95% CI [0.42, 1.66]). Unlike the results reported by Baker et al. (2002), peer-assisted learning within the same age range ($g = 0.14$; 95% CI [–0.09, 0.32]) and providing students feedback about their performance with goal setting ($g = 0.17$; 95% CI [–0.15, 0.49]) did not produce significant mean effects. Gersten, Chard, et al. (2009) also conducted regression analyses on 41 of the independent effects. Overall, the results indicated that mathematics interventions were generally effective across students, settings, and measures ($d = 0.63$, range = – 0.29 to 2.45, $p < .001$). The results from the regression analyses indicated that the majority of the contrasts did not result in effects significantly greater than the treatment average; however, there were a few exceptions. Interventions that incorporated the use of heuristics yielded significantly larger effects than the mean effect (1.21 effect increase, $p < .001$), as did explicit instruction as a teaching strategy (0.53 effect increase, $p < .05$). In contrast, when interventions utilized visual representations by teachers only (not by students), significantly smaller effects were reported (–0.68 effect decrease, $p < .10$).

Word problem solving. Other meta-analyses have looked exclusively at the effects of interventions that focused on word problem solving and computation. Xin and Jitendra (1999) searched relevant word problem solving literature from 1960 to 1996. Their meta-analysis included 14 group design and 12 single case design (SCD) interventions for students in first grade to beyond high school. Intervention participants had a documented disability or were considered at-risk for low achievement in mathematics because they received remedial mathematics instruction. The authors examined the effectiveness of word problem solving interventions regarding several variables including: participant characteristics, instructional features (e.g., duration, setting), methodological features, and skill maintenance and generalization. They also coded studies for instructional approaches in the intervention setting (representations, computer-assisted instruction, and strategy training including heuristics and direct instruction) and word problem task (one-step word problems, multi-step word problems, and mixed). Xin and Jitendra (1999) reported a large mean weighted effect size for group design studies ($d = 0.89$) and an average percentage of nonoverlapping data (PND) of 89% for SCD studies. Regarding the instructional approach, the largest effects were observed for interventions that used computer-assisted instruction ($d = 2.46$; 95% CI [1.97, 2.94]), representations ($d = 1.05$; 95% CI [0.79, 1.31]), and strategy instruction (d

= 1.01; 95% CI [0.85, 1.18]). Larger effects were reported for studies that included students at-risk for low achievement in mathematics (*d* = 2.22; 95% CI [1.88, 2.56]), compared to studies that only included students with documented disabilities (*d* = 0.68; 95% CI [0.55, 0.81]). Most studies addressed strategies for one-step word problems (*d* = 0.96; 95% CI [0.80, 1.13]) and a mix of word problems (*d* = 0.89; 95% CI [0.71, 1.08]).

Zhang and Xin (2012) conducted a follow up to the Xin and Jitendra (1999) synthesis; however, they included participants in kindergarten through 12th grade. Zhang and Xin (2012) reviewed 39 word-problem solving interventions published between 1996 and 2009, and included 10 SCD and 29 group design studies. The authors coded studies for several characteristics including participant risk status (disability, at-risk), intervention strategy (problem structure representation techniques, cognitive strategy training, computer-assisted instruction, traditional instruction), type of outcome measure used to the assess the effectiveness of the treatment (researcher-developed, high-stakes test, other standardized test), algebraic instruction (arithmetic, pre-algebraic), and problem tasks (simple problems, real-world problems).

Large mean effects were reported for group design studies (*d* = 1.85) and SCD studies (PND = 95%). For group design studies, larger effects sizes were associated with interventions that took place in an inclusive setting (*d* = 2.60; 95% CI [1.99, 3.21]) compared to a special education setting (*d* = 1.38; 95% CI [0.86, 1.83]). The intervention strategies coded yielded large effects, though there was some variability in the degree of the effects between problem structure representations (*d* = 2.64; 95% CI [1.96, 3.31]), cognitive strategy training (*d* = 1.86; 95% CI [1.07, 2.64]), and computer-assisted instruction (*d* = 1.22; 95% CI [0.62, 1.81]). Regarding the mathematics education variables, the authors reported no substantial differences between simple problems and real-world problems; however, arithmetic problem interventions yielded slightly larger effects (d = 2.01; 95% CI [1.51, 2.50]) than pre-algebraic problem interventions (M = 1.55; 95% CI [0.88, 2.23]). Finally, researcher-developed measures (d = 1.87; 95% CI [1.48, 2.26]) yielded larger effects than standardized measures (d = 0.60; 95% CI [–0.43, 1.63]). Due to the fact that most SCD studies had an average PND of 100%, no significant differences were reported between the moderators that were examined.

Computation. Codding et al. (2009) conducted a literature review of the effects of simple and moderate intensity computation interventions for participants who were identified as having difficulty in computation (i.e., low scores on an assessment or teacher referral). Simple interventions were identified as practices that improved students' performance but only required few changes in instruction; examples included reinforcement, feedback, changing practice opportunities, and goal setting. Moderate interventions were identified as instructional practices aimed at enhancing the existing mathematics instruction, such as peer-assisted learning, and cover-copy-compare computation instruction. The authors searched the relevant literature from 1980 to 2007, and included 12 group design and 25 SCD design studies. They evaluated the effect of the computation intervention regarding intervention characteristics, such as intensity of the intervention; the skills that the intervention was appropriate for; and the required training, materials, and support necessary for staff to implement the intervention. Furthermore, when it was possible the authors evaluated the range of effect sizes for different intervention characteristics according to computation outcome variables of rate (i.e., fluency of computation) and accuracy (i.e., correct performance not based on fluency).

The results were reported in the form of range of effect sizes yielded per study; meta-analytic techniques were not used. Codding et al. (2009) reported that features and types of effective computation interventions included goal setting, flash cards, taped problems, incremental rehearsal, and cover-copy-compare. Effect sizes ranged from negligible to large. For example, when considering moderate interventions, the effects of peer-tutoring ranged from 0.17 to 5.98 for rate as an outcome, and cover-copy-compare effects ranged from 0.75 to 6.29 for accuracy and from –0.34 to 7.72 for rate as an outcome. The effects of self-instruction ranged from 0.84 to 19.50 for accuracy and from 0.31 to 5.53 for rate an outcome. Regarding simple interventions, the effects of goal-setting ranged from zero to 1.20 for rate as an outcome and varied depending on the type of goal-setting (self-created goals, assigned goals, cooperative, and individual).

Reviews specific to elementary-aged students. While Baker et al. (2002) and Gersten, Chard, et al. (2009) reviewed mathematics interventions broadly for students through 12th grade, Kroesbergen and Van Luit (2003) focused their meta-analysis on younger children. The authors conducted a search of relevant literature from 1985 to 2000; they included 58 studies (61 interventions) for students with MD in kindergarten through a mean age of 12 years. MD referred to students at-risk for disabilities, low achievement in mathematics, or having a mathematics learning disability. The 61 interventions included 40 group design studies and 21 SCD studies.

In addition to coding for study (e.g., year of publication) and sample (e.g., mean age of participants) characteristics, the authors also coded specific mathematics content addressed by the intervention. Kroesbergen and Van Luit (2003) evaluated interventions in three different domains including preparatory mathematics (e.g., counting), basic skills (e.g., simple addition and subtraction), and problem solving strategies. They also examined the effect of the intervention in relation to participants' special needs status, study design, and use of peer tutoring. They reported moderate effects for group design studies (weighted $d = 0.62$) and large effects for SCD studies (weighted $d = 2.16$). Though each of the three mathematics domains yielded large effects, the authors reported larger effects for interventions that focused on basic facts ($d = 1.50$; $SD = 1.19$) versus interventions that focused on preparatory skills ($d = 0.92$; $SD = 0.72$) or problem solving skills ($d = 0.84$; $SD = 0.86$). Compared to other MD categories (d range = 0.83 to 1.09), the 23 interventions that included participants with learning disabilities produced larger effects ($d = 1.65$; $SD = 1.11$). Regarding the instructional format, interventions that included self-instruction yielded larger effects ($d = 1.77$; $SD = 1.12$) than interventions that included direct instruction ($d = 1.13$; $SD = 0.94$) or mediated instruction ($d = 0.52$; $SD = 0.97$). Kroesbergen and Van Luit (2003) also conducted a multi-level meta-regression and reported that four variables (design, duration, mathematics content domain, and instructional format) explained 69% of the variance in intervention effects.

Chodura et al. (2015) also reviewed mathematics interventions for students ages 6 to 12 years. Specifically, the authors included students with MD, including students with diagnosed dyscalculia and learning disabilities, and students who performed below the 26th percentile on a mathematics test or had mathematics achievement lagging more than 1 year behind grade-level expectations. The authors identified 35 group design studies that met inclusion criteria; the articles were published between 1985 and 2012. Chodura et al. (2015) examined variables such as instructional setting, level of mathematics competency addressed by the intervention content, and method of instruction. They reported a high significant mean standard difference ($\hat{\theta} = 0.83$; $SE = 0.13$; $p < .001$) across interventions. Significant effects were observed for interventions that were implemented in one-on-one settings ($p < .001$) and interventions that incorporated direct instruction ($p < .01$) or assisted instruction ($p < .05$). Similar to Kroesbergen and Van Luit's (2003) results, Chodura et al. (2015) reported a significant effect for interventions that focused on basic arithmetic competencies ($p < .05$).

Similar to Codding et al. (2009), Burns et al. (2010) reviewed the effects of computation interventions; however, the results of Burns et al. (2010) are limited to participants in second through sixth grade. The authors included 17 SCD studies (with a combined total of 55 participants) published between 1989 and 2007. Specifically, the authors examined the influence of collecting student data prior to the intervention to identify students or a target skill and compared the effects of acquisition and fluency interventions based on the baseline performance of participants (i.e., was the intervention at the instruction level or frustration level for the participant). Across the 17 studies, 10 interventions were in a cover-copy-compare format or used flashcard procedures; the remaining interventions included strategies such as self-monitoring and contingent reinforcement. The effects of the interventions ranged from negligible to moderate (phi range = 0.21 to 0.69). The authors reported no significant difference between studies that used assessment data collected prior to the intervention (phi = 0.52; 95% CI [0.34, 0.70]) and interventions that did not use assessment data (phi = 0.43; 95% CI [0.23, 0.63]). Further, they reported that acquisition interventions at the frustration level yielded significantly ($p < .01$) larger effects (phi = 0.84; 95% CI [0.76, 0.93]) than did acquisition interventions at the instructional level (phi = 0.49; 95% CI [0.29, 0.70]); there were too few effect sizes in fluency interventions at the instructional level to draw conclusions regarding the effect of baseline skill level.

Reviews of numeracy interventions. Unlike other areas of mathematics, there have been very few reviews of interventions that focus exclusively on content geared toward young students, such as early numeracy. Mononen et al. (2015) searched the literature on early numeracy interventions between 2000 and 2012 and identified 19 studies that examined the effects of numeracy interventions for participants between the ages of 4 and 7 years. Furthermore, participants must have presented with either low achievement in mathematics (i.e., performance below the 25th percentile on a mathematics measure) or were considered at-risk for academic failure by having low SES status. The authors examined intervention features such as setting, duration, numeracy content, required professional development and training for intervention agents, and instructional design. The results were presented as the proportion of studies that included specific intervention features and effects were not aggregated across studies (i.e., results did not include average effect sizes and authors did not conduct a meta-regression). The results of the review indicated that students in treatment groups made progress when the intervention included explicit instruction and peer-assisted learning strategies, and when interventions were presented as games, computer-assisted instruction, or in a concrete-representation-abstract (CRA) framework.

Malofeeva (2005) conducted a meta-analysis of interventions for preschool and kindergarten students (3 to 6 years old); the author searched the literature from 1977 to

2003 and identified 29 group design studies for inclusion. The author included students of varying ability levels (i.e., students with disabilities, at-risk for disabilities, and typically achieving); however, the initial mathematics performance of students was not taken into consideration for the analysis. This particular review was a dissertation; therefore, each of the included studies was extensively coded for several variables including, but not limited to, study and sample characteristics, instructional characteristics (e.g., duration, instructional approach, language of instruction, setting, materials), specific mathematics content (e.g., geometry, cardinality, computation, estimation), and how interventionists were trained. The results indicated a moderate mean treatment effect for mathematics interventions for young students ($g = 0.48$; 95% CI [0.35, 0.60]). Malofeeva reported larger effects for interventions that used a combination of guided and direct instruction approaches ($g = 0.63$; 95% CI [0.48, 0.78]) compared to interventions that only used guided instruction ($g = 0.43$; 95% CI [0.26, 0.59]) or only direct instruction ($g = 0.15$; 95% CI [–0.05, 0.35]). Studies that used standardized measures (i.e., distal measures) yielded larger effects ($g = 0.64$; 95% CI [0.46, 0.83]) than studies that used skill-specific measures ($g = 0.38$; 95% CI [0.22, 0.54]). Malofeeva (2005) also reported that the length of the intervention did not have a significant effect on the overall treatment effect, and the only specific mathematics skill that had a significant effect on the treatment outcome was measurement ($\beta = -0.36$, $p < .05$). Other skills that the author coded for interventions (e.g., counting, comparison, numeral identification) were not significant. Finally, Malofeeva examined the effect of specific instructional characteristics (e.g., feedback, reinforcement, elaboration, peer tutoring) and reported that significant predictors of the treatment outcome included one-to-one instruction, controlling task difficulty, group instruction, elaboration, and sequencing lessons.

Interventions specific to preschool. With regard to previous research on preschool mathematics interventions, the literature base is limited compared to research on mathematics interventions for students in kindergarten to 12th grade. To date, there is not a meta-analysis that solely evaluates the effect of mathematics interventions for preschool-aged participants. However, researchers have reviewed preschool interventions and specific curriculum in the Institute of Education Sciences (IES) Practice Guide (Frye et al., 2013) and in a national report on the effects of preschool curriculum on school readiness (Preschool Curriculum Evaluation Research Consortium [PCER], 2008).

The IES Practice Guide identified specific mathematics content areas that are critical to teach to young children and recommended effective instructional strategies for incorporating mathematics content into classrooms for children between the ages of 3 and 6 years. To prepare the Practice Guide, Frye et al. (2013) searched relevant literature from 1989 to 2011 and identified 29 studies that met What Works Clearinghouse (WWC) standards to be included in the report. These 29 studies were used to provide the panel's recommendations, and results were reported in the form of the level of evidence to support a recommendation (i.e., strong, moderate, and minimal evidence). Based on the literature, Frye et al. (2013) made five recommendations for teaching mathematics to young children including: (a) teach numbers and operations concepts, (b) teach geometry, measurement, and data analysis skills, (c) use progress monitoring tools, (d) teach students to view the world mathematically, and (e) integrate mathematics instruction throughout each school day. The authors did not report strong levels of evidence for any of the five recommendations; however, they did report a moderate level of supporting evidence for preschool and kindergarten instructional programs that focused on teaching students number and operations skills. In contrast, they found only minimal levels of evidence for each of the remaining recommendations (Frye et al., 2013).

The PCER initiative reviewed 14 preschool curricula, but the review did not explicitly focus on mathematics. The review was sponsored by IES as a method to conduct efficacy evaluations for widely used preschool curricula. The results of the report reflected individual evaluations of the different curricula (i.e., each curriculum was reviewed by a different research team); thus, different control conditions were used across reviews. The authors (PCER, 2008) examined the effect of each curriculum on students' reading and pre-reading abilities, early language, mathematics, and behavior. The review also examined classroom level outcomes such as teacher quality and teacher-child interactions. Only one of the 14 curricula had a positively significant impact on the overall mathematics composite during the preschool years ($p < .01$), *Pre-K Mathematics supplemented with DLM Early Childhood Express Math software*, and one curriculum had a significant impact on the WJ-Applied Problems subtest ($p < .01$), *DLM Early Childhood Express supplemented with Open Court Reading Pre-K.* During the kindergarten year, only *DLM Early Childhood Express supplemented with Open Court Reading Pre-K* continued to have a significant effect on WJ-Applied Problems. Though not to the level of statistical significance, five curricula in preschool and eight curricula in kindergarten produced negative effects on either the WJ-Applied Problems or on the overall mathematics composite.

Summary of Previous Syntheses of Mathematics Interventions

When considering all of the reviews discussed here, a few themes and inconsistencies emerge related to the effectiveness of mathematics interventions. Most authors coded for similar information including participant risk status, instructional strategies or approaches, and mathematics content. The results of several reviews revealed that interventions yielded larger effects when the following characteristics were present: explicit instruction, use of representations, use of strategy instruction, and computer-assisted learning strategies. Interventions that focused on basic arithmetic also consistently produced large effects. In contrast, inconsistent results were reported for interventions that employed peer-assisted learning strategies and provided students with feedback about their performance or required students to set goals about their performance. Mixed results were also reported for participant risk status and setting (i.e., inclusive versus special education). For example, some reviews reported that inclusive settings yielded larger effects, while other reviews found that one-to-one instruction yielded larger effects. Patterns in results are less clear specific to early numeracy interventions due to the limited number of reviews that have specifically focused on this age group in mathematics; however, the focus on teaching young students number and operations skills is supported in intervention research and in instructional recommendations (Frye et al., 2013; NCTM, 2006).

Limitations of Previous Syntheses

The current meta-analysis will address several limitations of previous reviews. First, the majority of reviews focused on students in kindergarten through 6th or 12th grade. Some reviews (Baker et al., 2002; Gersten, Chard, et al., 2009; Kroesbergen &

Van Luit, 2003) focused on identifying effective practices regarding mathematics interventions more broadly; for example, instructional techniques (e.g., direct instruction), instructional setting (e.g., small group), and intervention duration. Other researchers (Burns et al., 2010; Codding et al., 2009; Xin & Jitendra, 1999; Zhang & Xin, 2012) examined the effects of mathematics interventions that focused on a specific mathematics skill such as word problem solving and computation. Because of the broad age range of the participants in these reviews, researchers and practitioners may find it difficult to generalize findings to younger students. Furthermore, when a review includes many age groups, the content within each of the studies within the review is not always relevant to the instruction that preschool and early elementary students receive.

Second, only two reviews have specifically included studies of mathematics interventions for young children, and only one of these considered the numeracy content within the intervention (Malofeeva, 2005). Although Malofeeva (2005) coded intervention studies for numeracy content, the focus of this meta-analysis was not related to examining the effects of interventions according to the numeracy content. There are also limitations to this meta-analysis. The author did not consider the various levels of risk for MD of participants and examined the results of all intervention studies for participants equally (i.e., interventions that targeted students with disabilities or difficulties were examined with interventions that only included typically achieving students). Finally, Malofeeva's (2005) literature search ended in 2003, and since that date researchers have made many advances in developing and evaluating the implementation of early numeracy interventions. In contrast, Mononen et al. (2015) did consider the risk status of participants; however, the authors did not include any typically achieving students or studies that only included participants with disabilities. Furthermore, Mononen et al. (2015) did not aggregate effect sizes or report mean effects for specific intervention characteristics; the results were reported as the proportion of studies that had certain characteristics (e.g., results reported the proportion of studies that addressed number recognition skills). Based on the discussion of previous reviews of mathematics interventions and the limitations, the current meta-analysis will contribute to the literature in special education for students who struggle in mathematics in several ways.

Contributions of this Meta-Analysis

This meta-analysis is intended to make the following contributions:

1. This meta-analysis specifically examines the effect of early numeracy interventions, rather than mathematics interventions broadly. Skills in the numeracy domains are considered fundamental and necessary for learning more complex mathematics skills; therefore, the effects of numeracy interventions should be examined separately from other mathematics skills, such as computation and word problem solving. Findings from this meta-analysis may provide researchers and practitioners valuable information to develop effective interventions for young students and select important features to include in numeracy instruction.
2. A related contribution of this meta-analysis is that it quantifies the degree to which variability in treatment outcomes is related to participant risk status in mathematics. In this meta-analysis, I attempted to identify characteristics of effective interventions for all students, and determine if effective components differ based on risk status (i.e., disability or difficulty with mathematics). These

results are intended to contribute to guidelines for numeracy instruction and intervention for subgroups of students in the future.

3. The results of this meta-analysis provide an up-to-date review of numeracy interventions that includes both published and unpublished literature on numeracy interventions. This comprehensive approach provides a more thorough investigation of the overall effects of numeracy interventions and reduces publication bias.

Chapter III

Method

The purpose of this meta-analysis was to review intervention research on early numeracy instruction for preschool, kindergarten, and first-grade students. The specific goals were to determine the overall mean effect of early numeracy interventions and identify variables that explain the largest proportion of between-studies variance for the sample of studies. In this chapter, I describe the search process used to identify studies for inclusion, procedures to code each article and effect size, and the data analyses used to address each research question.

Search Process

To locate relevant studies for this meta-analysis, I conducted a systematic and exhaustive search process. This process was documented at each step. At each phase of the search and coding, reasons for study inclusion were documented as well as the number of studies excluded for each reason. An outline of the search process is described here.

1. I conducted an electronic search of literature on early numeracy interventions using Academic Search Premier, Education Source, ERIC, ProQuest Digital Dissertation, and PsycINFO databases from 1980 to June 2016. In an effort to obtain all relevant studies, the search was not limited to peer-reviewed journals, but instead also included conference proposals, technical reports, and research reports. The following combinations of search terms were used to search the databases (*math** OR *num**) AND (*intervention* OR *training* OR *remed** OR

instruction) AND ("*early intervention*" OR *preschool* OR *kindergarten* OR "*first grade*" OR "*early childhood*").

2. I conducted a search of reference lists of relevant studies, literature reviews, syntheses, and meta-analyses in mathematics instruction for students who are at-risk for mathematics difficulties or have disabilities (e.g., Chodura et al., 2015; Gersten, Chard, et al., 2009) and mathematics interventions for young students (e.g., Kroesbergen & Van Luit, 2003; Malofeeva, 2005; Mononen et al., 2015).
3. I searched the reference lists of relevant IES reports to identify additional articles and technical reports (Frye et al., 2013; Gersten, Beckmann, et al., 2009; Siegler et al., 2010).
4. I contacted authors who had published relevant studies to determine if they had any unpublished studies that met inclusion criteria for this meta-analysis.

Inclusion Criteria

Studies were included in this meta-analysis if the following inclusion criteria were met:

a) Studies were included if they examined the effects of school-based mathematics interventions, with the aim to identify interventions that can be used within multi-tiered systems of support. Specifically, for the purpose of this meta-analysis, an intervention was defined as a supplement to or supplant of general education mathematics curricula (core, Tier 1). Interventions reflected a change to typical instruction, such as varying instructional formats (e.g., one-on-one instruction), intensity (i.e., dosage at a rate prescribed by the intervention developer), and instruction was targeted to meet specific student weaknesses (Codding et al., 2009; Wanzek & Vaughn, 2007). In contrast, Tier 1 curricula are typically not designed to meet instructional needs of students who experience or are at-risk of experiencing difficulties with learning mathematics; therefore, studies that evaluated the effects of general education curricula were not considered for inclusion even if authors referred to the Tier 1 curriculum as an intervention (e.g., Chard et al., 2008; Clarke et al., 2011; DeLaoch, 2012; Jung, Hartman, Smith, & Wallace, 2013; Klein, Starkey, Clements, Sarama, & Iyer, 2008).

b) Studies were included if the authors evaluated the effectiveness of an academic intervention to improve early numeracy proficiency, and if interventions included at least one component from the early numeracy domains (Number, Relations, Operations). Thus, the treatment group must have received an intervention with more than 50% of the lessons or lesson content focused on one or more of the following early numeracy components: base-10 number systems, cardinality, composition, counting error identification, counting with one-to-one correspondence, decomposition, equivalence, magnitude comparison, missing number, number combinations, number line estimation, number line sequencing, numeral identification, ordinal numbers, set comparison, simple addition and subtraction with and without objects, subitizing, and verbal counting. Studies were excluded if the majority of the intervention content (more than 50%) was a different mathematics skill, such as computation fluency or geometry (e.g., Baroody, Eiland, Purpura, & Reid, 2012, 2013; Benson, 2013). This criterion was set to ensure that studies that focused on fluency building or other skills were not included in the meta-analysis simply because they included a review session on counting skills in the intervention, for example. Studies were also excluded if the focus of the intervention was on concepts of logical foundations or prerequisite skills such as number conservation, classification and seriation, and the oddity principle (e.g., Kidd, Pasnak, Gadzichowski, Ferral-Like, & Gallington, 2008; Pasnak, Holt, Campbell, & McCutcheon, 1991; Pasnak et al., 2009). These studies were not included because the focus of this meta-analysis was on academic interventions, whereas studies that focus on skills such as classification and seriation are typically categorized as cognitive interventions (e.g., training working memory; Kidd et al., 2008; Kroesbergen, van't Noordende, & Kolkman, 2014). Brief descriptions of the common academic components that make up the Number, Relations, and Operations domains are included in the section *Definition of Key Terms* (Chapter 1).

c) Studies were included if there was at least one treatment group and one comparison group, and the comparison group was representative of the same sample as the treatment group. Studies were excluded if they included students with chronic mathematics difficulty for the treatment group, but included only typically achieving students for the comparison group (e.g., Räsänen, Salminen, Wilson, Aunio, & Dehaene, 2009). For this reason, regression discontinuity designs based on mathematics screening cut-scores for participants were also not considered for inclusion (e.g., Bryant, Bryant, Gersten, Scammacca, & Chavez, 2008; Bryant, Bryant, Gersten, Scammacca, Funk, et al., 2008).

d) Studies were included if the participants were in preschool (at least 4 years old), kindergarten, or first grade. Since age and/or grade was a variable of interest in determining the effectiveness of the intervention, studies that included cross-grade level participants (e.g., included both kindergarten and first grade students) were included only if outcome data were disaggregated by grade or age. Children in preschool, kindergarten, and first grade were the focus of this meta-analysis because numeracy instruction typically occurs during these school years. Frye et al. (2013) identified moderate levels of evidence for teaching number and operation concepts to children 3 to 6 years old, and the CCSS emphasize numeracy concepts in kindergarten and first grade such as counting, cardinality, comparison, and simple arithmetic (NGA & CCSSO, 2010). Studies that included participants with a mean age less than 4 years old, students in second grade or above, and students across grade levels without disaggregated information by grade or age were excluded (Holmes & Dowker, 2013; Kaufmann, Handl, & Thöny, 2003; Whyte & Bull, 2008). Four years was selected as the cut-off for preschool participants because most states have state-supported preschool programs for 4-year-olds but very few children under the age of 4 years have access to public preschool programs (Barnett, Carolan, Squires, & Clarke-Brown, 2014). Children who have the opportunity to attend preschool at a younger age (i.e., 3 years) may represent a different population (e.g., higher family income) of students.

e) Studies were included if they employed an experimental or quasi-experimental design that included a control group and provided sufficient information (means and *SD*s or *F* statistics) to calculate effect sizes. Studies were excluded if effect sizes could not be calculated (e.g., Mulligan, Mitchelmore, Kemp, Marston, &

Highfield, 2008; Ramani & Siegler, 2011; Siegler & Ramani, 2008), and if they used single subject designs (e.g., Doabler et al., 2015; Murphy, Bates, & Anderson, 1984). Studies were also excluded if they only reported gain scores (WWC, 2014; e.g., Young-Loveridge, 2004).

f) Studies were included if they included at least one mathematics outcome measure to evaluate the effect of the intervention.

g) Studies were included if they were published in English.

Search Coding Procedures

Each study retrieved during the search process was coded in a two-phase process to determine eligibility for inclusion. In total, 2,844 studies (dissertations, peer-reviewed manuscripts, conference presentations, research reports) were identified for review in Phase 1, which included a review of study titles, keywords, and abstracts. In Phase 1, approximately 94% of articles were excluded for the following reasons: irrelevant focus (53.5%), intervention that was not mathematics related (26.5%), focused on numeracy but was not an intervention (8.6%), duplicate across databases that was not immediately deleted from the search (2.3%), included only participants with severe disabilities (1.5%), single subject design (.5%), and the participants were the wrong age (< 1%). Approximately five percent (*n* = 163) of studies were identified for review in Phase 2.

If the title, keywords, and/or abstract of the study were relevant to the purpose of this study, the full text of the article was reviewed to determine if the study met the full inclusion criteria. Studies were excluded if they focused on evaluating the effects of a comprehensive preschool curriculum (i.e., curriculum for reading, mathematics, social skills) or a Tier 1 mathematics program (22.7%) or if they focused on evaluating the effects of an intervention but the content was not numeracy-based (19.6%). Studies were also excluded if the necessary information to calculate effect sizes was not provided (including no mathematics outcome measure given; 11.0%). Other studies were excluded because the study design was single subject (7.4%), participants were not the target age (3.7%), duplicate studies with other publication types such as conference presentations (3.1%), and for other reasons (12.1%; the study was not an intervention, gifted education program, control group represented a different population of students, commentary). Finally, one study (Gersten et al., 2015) was excluded because authors imputed post-test data for approximately 11% of the sample in order to conduct an intention-to-treat (ITT) analysis. With ITT, there is the potential that participants who did not actually receive the intervention be included as a subject in the analysis as receiving the intervention. For example, if a student is placed into the treatment condition and drops out of the study, they will be included in all analyses (Gupta, 2011). ITT also preserves the sample size; therefore, Gersten et al. (2015) used multiple imputations to estimate the outcome data for 113 participants. This study was perceived as an outlier and different from the rest of the sample of included studies, as no other study in the sample imputed missing post-test data.

Coding Studies Procedures

The studies that met inclusion criteria were coded with regard to the following: study information, methodological characteristics, participant characteristics, intervention characteristics, outcome measures, effect size, and confidence of codes. Appendix A includes all of the variables that were coded, definitions of the variables, and specific information regarding how variables were coded.

Study information. Variables related to the study were coded, including: year of publication, country in which the study was conducted, and type of publication (e.g., peer-reviewed journal, dissertation).

Study quality. The quality of the study was coded on a 0- to 5-point scale regarding the methodology and quality of information provided for key variables. The variables included: disaggregated demographic information for treatment and control groups (i.e., demographic information was reported for participants separately for each treatment and control group), demographic information for intervention agents to allow for replication (e.g., gender, level of education), fidelity of the treatment condition, description of the control condition, and attrition information.

Methodological characteristics. Variables related to the study methodology were coded. These variables included total sample size, sample size of the treatment and control groups, attrition of the total sample, treatment, and control groups expressed as percentages, assignment to the treatment condition (i.e., random assignment, nonrandom assignment, matching), nature of the treatment condition (i.e., did the intervention supplant or supplement the core mathematics instruction), nature of the control condition (e.g., business as usual, other mathematics intervention). Study design (e.g., post-test only, pre- and post-test), and information regarding the independence of effect sizes from each study (i.e., if there was more than one treatment group) were also coded.

Participant characteristics. Variables related to the participants' demographics were coded, and included mean age and grade level (preschool, kindergarten, first grade), gender, race/ethnicity, English Learners (ELs), free/reduced lunch (FRL), disability or at-risk status. When necessary, demographics reported for each independent effect size were converted from percentages to numbers as a method of accurately reporting the total number of participants in each demographic group. Criteria and specific measures used to determine at-risk status were recorded when studies identified participants as having mathematics difficulty (MD). Regarding MD, studies were coded for the following information: participants had a documented disability, participants were considered at-risk for MD, and study did not report inclusion criteria for participants.

Regarding grade level of participants, studies conducted outside of the U.S. were coded according to participant age and year in kindergarten. For example, some countries require two years of kindergarten. Although both years are referred to as “kindergarten” or “preschool,” if participants were in their first of two years of kindergarten, the study was assigned the code of “preschool” and if participants were in their second of two years of kindergarten, the study was assigned the code of “kindergarten.” This change in coding of grade level (two studies had grade levels changed) was also checked in relation to the age of participants to ensure alignment of the codes. This coding allowed for closer alignment in age for all studies coded as “preschool” and all studies coded as “kindergarten” in order to generalize results to specific age groups.

Intervention characteristics. Characteristics related to the intervention were also coded. These variables included instructional arrangement (e.g., individual, small group, peer-assisted, whole-class, one-to-one), intervention agent (i.e., who implemented the intervention), and intensity of the intervention (number of sessions, min/session, and duration). To capture what the instruction encompassed and what instructional strategies participants were exposed to, specific characteristics of instruction were also coded (e.g., use of manipulatives, guided practice, game-based format). Finally, the interventions were coded for the specific skills addressed in each of the numeracy domains (Number, Relations, and Operations). Interventions that contained at least one skill in a particular domain were identified as addressing that numeracy domain; for example, interventions were not required to address *all* skills within the Number domain for the intervention to receive a code for Numbers skills. Explicit information from the study, appendices with lesson information, or references to previous work on a similar intervention were used to determine which numeracy domains and skills were included in each intervention.

Dependent measures. Although all outcome measures were recorded, only the mathematics measures were coded extensively. For example, outcome measures that measured attention and motivation were not considered in this study. A description of each mathematics outcome measure was recorded; specifically, the early numeracy skills that the measure assessed were recorded. Other characteristics of measures that were coded included reliability coefficients and type of reported reliability, validity coefficients and source of validity. Mathematics measures were coded as either broad (e.g., achievement test, comprehensive measures encompassing many grade-level mathematics skills) or narrow (e.g., tests that assessed single skills such as simple addition or numeral identification), and proximal (test material aligned closely to intervention content) or distal (test material was not specifically aligned to the intervention material). Measures were also coded as either norm-referenced, researcher developed, or unknown.

If only one mathematics measure was used in the study, this measure served as the primary outcome measure by default; however, many studies included multiple mathematics outcome measures. When researchers used more than one mathematics measure, the measures were not averaged to report a composite effect size for the treatment effect because averaged effect sizes based on multiple outcome measures may mask the unique results observed for specific early numeracy outcome measures. Furthermore, an average effect size may represent effects in other areas of mathematics such as geometry, measurement, or computation fluency that are not the focus of this study. Furthermore, if more than one outcome measure is taken into consideration to represent the effects of an intervention, those effect sizes are dependent because they were sampled from the same group of students. To correct for this dependency, correlations between the outcomes should be considered and most researchers do not provide the necessary data to determine the correlations between outcomes (Borenstein et al., 2009; Cooper et al., 2009). Therefore, I made a conscious effort to select one mathematics outcome measure that was the best representation of the construct of early numeracy and the results of only one outcome measure were used to calculate the effect size. The following hierarchical process was used to select the primary measure to calculate the effect size to represent the treatment effect.

1. Broad mathematics measures that were closely aligned to the intervention were selected as the primary outcome measure if available (e.g., *Number Sense Brief*; Jordan, Glutting, Ramineni, & Watkins, 2010). The purpose of this meta-analysis was to determine the effect of early numeracy interventions on mathematics performance, and broad numeracy measures were considered more representative of the construct as a whole, compared to narrow measures (e.g., 1 min timed measure of oral counting). Measures that were closely aligned (proximal) were considered more relevant than distal measures.

2. If a broad proximal mathematics measure was not available, the next optimal choice was a broad distal measure of mathematics performance (e.g., *Test of Early Mathematics Ability* – 3rd edition; Ginsburg & Baroody, 2003). Broad distal measures were considered more optimal than either narrow proximal or narrow distal measures because the reported effect size would indicate the treatment effect on students' overall mathematics performance, as opposed to one single skill.
3. If neither type of broad mathematics measure was used in the study, narrow proximal mathematics measures were selected as the primary outcome. Proximal measures were selected over distal measures in order to provide the most accurate effect size as related to the intervention content.
4. Finally, narrow distal measures were selected as the primary mathematics outcome to determine the treatment effect if no other mathematics measures were used.

Finally, if a study used a delayed posttest or maintenance test, those results were also recorded, along with the length of time (number of weeks or months) elapsed between the immediate posttest and maintenance test. Delayed and maintenance effect size were calculated separately from the immediate posttest treatment effects.

Confidence codes. A 3-point scale for the confidence in codes recorded for accuracy of information reported in studies was also assigned (Kazdin, 1977; Orwin & Cordray, 1985). For example, some studies reported information in one part of the study (e.g., in text description of the sample size and demographics) that did not match the information reported in another part of the study (e.g., sample size and demographics reported in tables). This code represents the confidence I had in the codes I assigned to each treatment groups; one of the following codes was applied to each treatment group:

- *Very confident:* This code represented that I was "certain or almost certain" about the codes recorded in the coding sheet (Orwin & Cordray, 1985). In other words, all or nearly all of the information reported in the study was clear and accurate. This code was applied to studies that had 0-1 discrepancies.
- *Mostly confident:* This code represented that I was certain or almost certain about accuracy of the majority of information presented in the study. In other words, there were two, three, or four perceived errors (discrepancies) in what the authors reported.
- *Somewhat confident:* This code represents that several of the codes and information reported in the study were largely in question. The study contained five or more perceived errors (discrepancies) in what the authors reported.

In addition to the code given to each study, notes regarding the discrepancies also were recorded (e.g., demographics in the table did not match the text, sample size in the abstract did not match table).

Effect size confidence. Because the effect size is the common metric in a meta-analysis it is important to acknowledge the accuracy of the effect size reported. Sometimes, data from the original study were converted or reconfigured in order to compute the effect size of interest in this meta-analysis. A 3-point scale for the confidence in the effect size calculation was also recorded (Borenstein et al., 2009):

- *Highly estimated:* the information available to calculate effect size was limited (e.g., N and p-value only).

- *Some estimation:* complete statistics were reported but there was some question regarding the precision or accuracy (e.g., unconventional statistics were reported that required conversion or use of significance tests that required conversion, accuracy of the sample size was questionable).
- *No estimation:* all conventional statistical data were present to calculate effect size (e.g., means, *SD*s).

Interrater Agreement

I served as the first rater for all codes, and a second rater, who was a doctoral student in educational psychology, independently coded 13 studies (39%) that I selected using a random number generator. I trained the second rater to use the coding rubric, and the second rater was required to reach 90% agreement on practice articles prior to coding the articles in this study. Articles that did not meet inclusion criteria were used for training purposes. Interrater agreement was calculated as: [agreements ÷ (agreements + disagreements) × 100]. We discussed any coding discrepancies to determine the final code to be used in the analyses.

Due to the large number of missing data on participants' demographics (e.g., gender, ELs, FRL, race/ethnicity), demographic codes were not considered as part of the calculated inter-rater reliability. The large number of agreements in demographics (due to more than 50% of codes being a "0" or "not reported" by both coders) would have inflated the inter-rater reliability estimate; therefore, only codes related to study characteristics, study quality, methodology, and intervention characteristics were included in the calculation of interrater reliability. However, the second rater did code demographics in order to spot-check any disagreements. The mean interrater reliability across categories was 90%; treatment characteristic codes (i.e., features of instruction and mathematics content) had an average agreement of 87%, study quality codes had an average agreement of 92%, intervention feature codes had an average agreement of 92%, methodology codes (e.g., assignment to treatment) had an average agreement of 94%, and study information had an average agreement of 100%.

In addition to double-coding 39% of all studies, the second rater also checked the first author's entries into the database for analyses (e.g., second author checked all inputted means and *SD*s used to calculate effect sizes in the database by cross-referencing all entries with the manuscripts). Discrepancies in the database were discussed and resolved prior to data analysis.

Data Analysis

An effect size allows readers and researchers to evaluate studies on a common scale according to magnitude and direction. In intervention research, the effect size represents the impact of the treatment (i.e., the degree of success of the intervention; Borenstein et al., 2009). The following sections describe the analyses that I conducted using Comprehensive Meta-Analysis (CMA) and IBM SPSS software.

Descriptive results. Before conducting analyses for each of the research questions, I summarized descriptive results for the interventions included in this study. Specifically, I provided aggregated results for each category in which studies were coded including study information (e.g., mean publication date), methodological characteristics (e.g., how many studies used random assignment), participant characteristics (e.g., the number of studies that included participants with MD), intervention characteristics (e.g., the average intervention duration), dependent measures (e.g., the number of measures administered per study), and confidence codes (e.g., total scores). This information was compiled to give readers a clear picture of the sample of interventions included in this study.

Meta-analysis results. First, in conducting data analyses to answer research questions 1, 2 and 3, I identified potential outliers as those effect sizes that were more than 3.0 *SDs* above the random weighted mean effect size (Cooper et al., 2009). Results for research questions 1, 2, and 3 are presented with and without the inclusion of the outliers. Then, based on the results of the removal of outliers for research questions 1, 2, and 3, I determined that the outliers would be removed from the regression analysis to answer research question 4. Data analyses for all research questions are discussed below.

Calculation of effect sizes. Based on the WWC guidelines (WWC, 2014), effect sizes were calculated as unbiased Hedges' *g* (Hedges & Olkin, 1985). Hedges' *g* accounts for small sample sizes across studies and uses the following correction factor (*J*) (Cooper et al., 2009):

$$J(df) = 1 - \frac{3}{4\,df - 9}$$

For studies that used a post-test only design, effect sizes were calculated by dividing the difference between the scores for the treatment and control groups at post-test by the pooled standard deviation. In other words, where Y_T represents the treatment group mean, and Y_C represents the control group mean, and S_{within} represents the within-groups standard deviation pooled across groups:

$$d = \frac{Y_T - Y_C}{S_{within}} \qquad S_{within} = \sqrt{\frac{(n_1 - 1)\,S_1^2 + (n_2 - 1)\,S_2^2}{n_1 + n_2 - 2}}$$

Regarding S_{within}, n_1 and n_2 represent the sample sizes of the treatment and control groups, and S_1 and S_2 are the standard deviations in the two groups.

If studies with a pre-post design did not report adjusted post-test means for non-equivalence of groups at pre-test, effect sizes were calculated as the "difference between the mean outcome for the intervention group and the mean outcome for the comparison group, divided by the pooled within-group standard deviation of the outcome measure" (WWC, 2014, p.22). In other words, where $Y_{1,2Post}$ represents the unadjusted post-test sample means for the treatment and control groups, and $Y_{1,2Pre}$ represents the unadjusted pre-test sample means for the treatment and control groups, and S_{within} is the within-groups standard deviation pooled across groups (Cooper et al., 2009):

$$d = \frac{(Y_{1Post} - Y_{1Pre}) - (Y_{2Post} - Y_{2Pre})}{S_{within}}$$

When studies did report adjusted posttest scores, effect sizes were calculated as the mean standardized difference between the experimental and control condition divided by the pooled standard deviation at posttest. Specifically, the following equation was used, where $\bar{Y}_{1Adjusted}$ and $\bar{Y}_{2Adjusted}$ represent the adjusted posttest means for the treatment and control groups, and $S_{withinAdjusted}$ represents the reported adjusted within-groups standard deviation pooled across groups (Cooper et al., 2009):

$$d = \frac{\bar{Y}_{1Adjusted} - \bar{Y}_{2Adjusted}}{S_{withinAdjusted}}$$

When a study reported the results of the intervention as a *t*-test value (*t*) without reporting the means or *SDs* for the outcome measures, formulas for conversion were applied, where *df* = degrees of freedom:

$$d = \frac{2t}{\sqrt{df}}$$

Finally, when the treatment effect was calculated (*d*), the following equation was used to determine *g*:

$$g = J(df)d$$

Effect sizes were interpreted in the following manner: 0.20 = small, 0.50 = moderate, and 0.80 or greater = large (Cohen, 1988).

Sensitivity analysis. From a preliminary examination of studies that were likely to meet inclusion criteria, I recognized that some studies contained one control group and two or more intervention groups that both focused on early numeracy skills (e.g., treatment group A received a counting intervention and treatment group B received a comparing sets intervention). The effect sizes calculated from the difference in means between the treatment group A and control and treatment group B and control are not independent of each other. Although there is not an agreed upon approach to addressing the issue of dependent groups in the same meta-analysis, the approach that I used was to include both treatment A versus control and treatment B versus control (which ignores the issue of dependency), and conduct a sensitivity analysis for (a) the overall treatment effect, and (b) the final meta-regression model. A sensitivity analysis helps evaluate how results could have been different if only one of the treatment groups was compared to control to determine if including dependent treatment groups altered the results of the study. When coding the studies that included more than one treatment group and only one comparison (control) group, I identified the "selected" and "non-selected" treatment groups. Selected treatment groups represented the group that received an intervention that was most closely aligned to the early numeracy skills that were coded for this meta-analysis. A summary of the studies with dependent effects, the selected intervention group, and a brief description of the reason for selected is presented in Appendix B. I present the results from the sensitivity analysis as: (a) using only “selected” treatment groups versus control groups, and (b) using only “non-selected” treatment groups versus control groups. I compare findings for the overall treatment effect and the final meta-regression model to determine if the results of the sensitivity analyses converge.

Estimates of effect sizes. As discussed in the Definition of Key Terms section, a fixed-effect model assumes that there is one true effect size that represents all studies in the analysis and observed differences in the effect size estimate are attributed to sampling error. In contrast, in a random-effects model the true effect may vary. Variation in effect sizes in a random-effects model is attributed to variables in addition to sampling error (Borenstein et al., 2009).

For this study, I used a random-effects model because this model assumes that instead of having one true effect size that represents all numeracy interventions, there is actually a distribution of true effect sizes. The combined effect represents the mean of the population of true effects (Borenstein et al., 2009). The studies included in this meta-analysis represent the work of many researchers and research teams who have operated independently. Furthermore, although the interventions were similar in that they represented the same type of early numeracy intervention, they differed on important aspects that may cause variation in effect sizes such as duration, training, and materials. It was highly unlikely that each of the interventions were equivalent in all features, as well as the populations they served; therefore, a random effects model was used. In the following sections, I discuss the specific data analyses used to answer each research question.

Publication bias. Although a meta-analysis provides an accurate synthesis of the studies included in the analysis, the mean treatment effect reported in the results may be biased if the sample of studies is biased. For example, studies that reported significant and positive results are more likely to be published than studies that report null or negative results (Borenstein et al., 2009; Franco, Malhotra, & Simonovits, 2014); therefore, studies included in a meta-analysis may actually overestimate the true treatment effect. To address publication bias, I took several steps. First, I conducted a comprehensive review of published and unpublished literature, including research reports, conference proposals and papers, and dissertations. Second, I generated a funnel plot to examine the symmetry of treatment effects around the underlying true effect. A funnel plot is an illustrative example of publication bias that plots effect size on the *X* axis and sample size or standard error on the *Y* axis. Studies with larger samples appear toward the top of the plot and studies with smaller samples appear toward the bottom of the plot. The studies are distributed symmetrically about the mean effect size on a funnel plot in the absence of publication bias. If publication bias is present, a gap would exist in a corner of the plot; for example, a gap in the left corner of the plot would suggest that studies with smaller effect size values are "missing" from the sample (Borenstein et al., 2009). Finally, I conducted the classic *Fail-safe N* analysis. The *Fail-safe N* analysis estimates the number of missing studies that would need to be incorporate in the analysis before the mean effect size became nonsignificant (Borenstein et al., 2009). If the number of studies needed to nullify the effect is small, there would be concern of publication bias.

*Research question 1***.** To answer the first research question, *what is the overall mean effect of early numeracy interventions on mathematics outcomes for students in preschool and early elementary, and how variable are the effects?*, I calculated a mean effect size (weighted by the inverse of the effect size variance), as well as the 95% confidence interval of the average treatment effect. Weighting studies allowed for studies with a larger sample size to carry more weight, as the treatment effects from those studies are more precise. I also conducted an overall Q test (i.e., test of homogeneity) to examine variation among the distribution of effect sizes. A statistically significant Q statistic represents a heterogeneous distribution of treatment effects, and supports further examination of potential causes of the variation, which is addressed in subsequent research questions.

In addition, to supplement the Q test, I also calculated I^2 to determine the proportion of variability in effect sizes that was due to heterogeneity rather than sampling error (Borenstein et al., 2009; Cooper et al., 2009). The larger the proportion, the more variability that is attributed to *true* differences among treatment effects, compared to differences due to sampling error.

Research question 2. To answer the second research question, *which early numeracy domain was most investigated, and which domain produced the largest effect size?*, I calculated general frequency counts of how many times each early numeracy domain (Number, Relations, and Operations) was addressed in the corpus of included articles. I also determined the average weighted effect for each domain (using the same procedures described above for calculating effect sizes) and calculated the 95% confidence interval for the average treatment effect.

Research question 3. To answer the third research question, *what were the differential treatment effects of early numeracy interventions on mathematics outcomes across study characteristics, participant characteristics, and intervention characteristics?* and examine the potential causes in variation of the overall mean effect, I conducted a series of analyses. First, I calculated the weighted average effect sizes for appropriate subcategories (e.g., participants with MD versus participants without MD), and calculated 95% confidence intervals for those average treatment effects. Then, I examined variation within and across subcategories with *Q* tests. Generally, I replicated the analyses from research questions 1 and 2, but used the analyses to compare subcategories regarding average weighted effects and tests of homogeneity.

Research question 4. To answer the research question, *which variables (e.g., age, instructional format) accounted for the most between-studies variance for the total sample, and for the three domains separately?* and to explore the potential sources of variation in treatment effects, I conducted a meta-regression. When I^2 is large it may be helpful to conduct a meta-regression to explain the source(s) of variation across the treatment effects. In a meta-regression, the independent variables are at the study/independent group level instead of the subject level, whereas the dependent variable is the treatment effect that represents the independent group (Borenstein et al., 2009). As an attempt to explore the sources of variation in the total sample and the sample of studies representing each domain separately, I considered several factors. The results of the research questions 1, 2, and 3 and previous research on the effectiveness of specific instructional features helped inform which variables were included in the final meta-regression. I also considered which variables were malleable attributes (or features that can be controlled) that practitioners and researchers could manipulate in future studies or in practice to replicate effective interventions. For example, the year a study was published was not considered as a variable to include in the meta-regression, because although study year may provide some insight into the context of the intervention, it does not provide any practical implications for teachers if found to be a significant source of between-groups variation.

In order to explore the potential sources of variation in each of the early numeracy domains separately, I took a slightly different approach in determine the final models for the meta-regression. When considering the sample of studies from each of the three early numeracy domains separately, I considered a restricted set of variables related to the early numeracy skills. For the set of Number interventions, I only considered the Number skills as potential variables related to mathematics performance and I did not include Relations or Operations skills in the final model; similarly, for the set of Relations interventions, I only considered the Relations skills as potential variables related to mathematics performance and I did not include Number or Operations skills; finally, for the set of Operations interventions, I only considered the Operations skills as potential variables related to mathematics performance and I did not include Number or Relations skills in the final model. I explored the construction of the models in this way because I was interested in determining which early numeracy skills explained variation in treatment effects in a restricted sample (i.e., which Number skills account for variation in Number skills intervention). With these models, malleable factors as well as results from the previous research questions also helped drive the creation of the final model. In order to consider all Number domain skills in the Number model, however, each Number skill was added into the model and non-significant Number skills were removed in an iterative process until the most parsimonious model was identified. The same process was used for Relations and Operations final models.

I examined which variables had sources of between-groups variation and accounted for the variance in the average effect size. In primary studies, R^2 is reported to represent the proportion of observed variance and each observation is given the same weight, but in a weighted least squares meta-regression, each effect size is given a different weight and the computation is based on true variance (Willett & Singer, 1988). Therefore, I reported pseudo R^2 for the models presented for the total sample and the three domains separately. Pseudo R^2 was calculated as:

$$\text{Pseudo } R^2 = \frac{Tau^2_{explained}}{Tau^2_{total}}$$

The following random effects model represents a meta-regression model. T_i represents a calculated effect size estimate of the true effect size θ_i (standardized mean difference) for each of k independent effect sizes, $i = 1, \ldots, k$. The variance, v_i, is in the form of $1/n_i$, where n_i is the sample size of effect size i. I assume that the error e_i of the estimation of the effect is independent and with mean of zero and a variance of v_i. So, (Cooper et al., 2009):

$$T_i = \theta_i + e_i$$

Regarding the prediction model based on the study characteristics that I choose to put into the model (Cooper et al., 2009):

$$T_i = \beta_0 + \beta_1 X_{i1} + \beta_1 X_{i2} + \ldots + \beta_p X_{ip} + u_i + e_i$$

Where (Cooper et al., 2009):

β_0 represents the model intercept;

$X_{i1}, \ldots, X_{ip}$ represents coded characteristics to predict θ_i;

$\beta_1, \ldots, \beta_p$ represents the regression coefficients;

u_i represents the random effect of i, or the deviation of effect size i's true effect from the value predicted from the model. I assume that each random effect, u_i, is independent and a mean of zero and a variance of σ_θ^2.

So, $u_i + e_i$ will represent the total variance in the observed effect size, T_i.

Chapter IV

Results

In this chapter, I provide descriptive results for each category of coding conducted for this meta-analysis, including study information, methodological characteristics, participant characteristics, intervention characteristics, dependent measures, and confidence codes. Then, I provide detailed results for each of the research questions guiding this study; meta-analysis results are presented with and without potential outliers.

Descriptive Results

Of the 163 studies identified for full review, 33 met inclusion criteria for this meta-analysis (21%), with a total of 51 treatment groups. Table 1 provides basic study information for each of the included studies (and separate treatment groups when applicable), including grade, sample size, type of interventions, participant risk status and criteria for risk, primary outcome measure and treatment effect (Hedges' *g*). Note that some results are reported related to the study level characteristics (n = 33), and some results are reported related to the treatment group (k = 51).

Study information. The studies were published between 1984 and 2016 (*M* publication date = 2009; *SD* = 7.72 years); 55% were published within the last 5 years. One study was a doctoral dissertation, and the remaining studies were published in peer-reviewed journals (at the time of the literature search one study was under review for a peer-reviewed journal, but was published at the time of analyses). The majority (n = 20; 61%) of studies were conducted in the U.S. Other studies were conducted in the Netherlands (n = 5) and Finland (n = 2), and six studies were each conducted in one of the following countries: Austria, Belgium, France, Germany, India, and Italy.

Table 1
Study and Intervention Information

Study (year)	G[a]	*N*	Treatment	Group (risk)[b]	Criteria for Risk	Outcome Measure	Hedge's *g*
Bryant et al. (2011)	1	204	Number, operations, problem solving	MD	Lowest 35% of the sample on TEMI-PM	TEMI-PM	0.56
Clarke, Doabler, Smolkowski, Baker, et al. (2016)	K	126	Whole number procedural fluency and conceptual understanding (ROOTS)	MD	Teacher nomination; of students nominated, 91% scored <10th pc. TEMA	TEMA	0.38
Clarke, Doabler, Smolkowski, Kurtz Nelson, et al. (2016)	K	203	ROOTS	MD	Lowest 10-12 students per class on composite score for NSB and ASPENS	RAENS	0.71
Clarke et al. (2014)	1	77	Number sense, number combinations, place value, computation, word-problem solving	MD	Lowest 10 students per school on Missing Number and Quantity Discrimination	ProFusion	0.84
Clements (1984)	P	45	Counting, cardinality	TA		Number Foundations	4.31
Codding, Chan-Iannetta, George, Ferreira, & Volpe (2011)	K	65	Number concepts, number comparisons, adding and subtracting concepts (Peer-Assisted Learning Strategies; [K]PALS)	TA		TEMA-3	0.40

Table 1 continued…

Study (year)	G[a]	*N*	Treatment	Group (risk)[b]	Criteria for Risk	Outcome Measure	Hedge's *g*
Desoete & Praet (2013)	K	132					
Group 1a			Computer-assisted instruction (CAI) on *counting* and cardinality	TA		TEDI-MATH	1.41
Group 2a			CAI on *counting* and cardinality	MD	<25th pc. TEDI-MATH	TEDI-MATH	1.64
Group 1b			CAI on *comparing* quantities, number words, numerals	TA		TEDI-MATH	0.65
Group 2b			CAI on *comparing* quantities, number words, numerals	MD	<25th pc. TEDI-MATH	TEDI-MATH	1.01
Doabler et al. (2016)	K	292	ROOTS	MD	Lowest 10 students per classroom on composite score for NSB and ASPENS	RAENS	1.09
Dyson, Jordan, Beliakoff, & Hassinger-Das (2015)	K	126	Whole number concepts related to counting, comparing, and				
Group 1			manipulating sets with *number sequence* practice	MD	<25th pc. NSB	NSB	0.34

Table 1 continued…

Study (year)	G[a]	*N*	Treatment	Group (risk)[b]	Criteria for Risk	Outcome Measure	Hedge's *g*
Group 2			Whole number concepts related to counting, comparing, and manipulating sets with *fact retrieval* practice	MD	<25th pc. NSB	NSB	0.87
Dyson, Jordan, & Glutting (2013)	K	121	Whole number concepts related to counting, comparing, and manipulating sets	AR	Low-income; average school free/reduced lunch = 91%	NSB	0.64
Fuchs et al. (2005)	1	564	Number concepts, sequencing numbers, place value, basic facts	MD	Lowest students identified with CBM battery, then 4 weeks of progress monitoring to identify lowest 21% of students	First grade concepts and applications	0.56
Fuchs, Fuchs, & Karns (2001)	K	162					
Group 1			KPALS	HA	Performance >1.5 *SD*s above the mean on SESAT	SESAT	-0.16
Group 2			KPALS	TA	Performance within .75 *SD*s of the mean on SESAT	SESAT	0.44
Group 3			KPALS	MD	Performance >1.5 *SD*s below the mean on SESAT	SESAT	0.38

Table 1 continued…

Study (year)	G[a]	*N*	Treatment	Group (risk)[b]	Criteria for Risk	Outcome Measure	Hedge's *g*
Group 4			KPALS	DIS	Identified or referred for special education	SESAT	0.43
Fuchs, Fuchs, Yazdian, & Powell (2002)	1	325					
Group 1			PALS	HA	Teacher identified level of proficiency	SAT (aligned items)	0.15
Group 2			PALS	TA	Teacher identified level of proficiency	SAT (aligned items)	0.16
Group 3			PALS	MD	Teacher identified level of proficiency	SAT (aligned items)	0.19
Fuchs et al. (2013)	1	591					
Group 1			Basic knowledge, relations, and *speeded* practice for fact retrieval	MD	Latent class analysis to determine MD risk	Word Problems	0.23
Group 2			Basic knowledge, relations, and *non-speeded* practice for reinforcing relations skills	MD	Latent class analysis to determine MD risk	Word Problems	0.28
Hansmann (2013)	1	123					
Group 1			Taped *traditional* quantity discrimination, students chose larger of two numbers	TA		Number Knowledge Test	0.00

Table 1 continued…

Study (year)	G[a]	*N*	Treatment	Group (risk)[b]	Criteria for Risk	Outcome Measure	Hedge's *g*
Group 2			Taped *triangle* quantity discrimination, students chose which of two numbers was "closest" to a number	TA		Number Knowledge Test	0.10
Hassinger-Das, Jordan, & Dyson (2015)	K	124					
Group 1			Counting, number relations and operations	MD	Performance <25th pc. NSB	NSB	0.34
Group 2			*Story-book* activities on counting, number relations and operations	MD	Performance <25th pc. NSB	NSB	0.21
Jordan, Glutting, Dyson, Hassinger-Das, & Irwin (2012)	K	128	Whole number concepts related to counting, comparing, and manipulating sets	AR	Low-income; average school free/reduced lunch = 93%	NSB	1.10
Kaufmann Delazer, Pohl, Semenza, & Dowker, (2005)	K	34	Inquiry-based focus on teaching conceptual skills of number and counting	TA		Calculation Battery	0.79
Kyttälä, Kanerva, & Kroesbergen (2015)	P	38	Counting, matching numerals and quantities, number sequences	TA		ENT (Finnish edition; counting subtest only)	0.50
Mohanty & Mishra (1994)	P	30	Counting, cardinality	TA		Number Concepts	6.63

Table 1 continued…

Study (year)	G[a]	*N*	Treatment	Group (risk)[b]	Criteria for Risk	Outcome Measure	Hedge's *g*
Obersteiner Reiss, & Ufer (2013)	1	147					
Group 1			CAI on *exact* processing of numbers and quantities	TA		Hamburger Rechentest	0.45
Group 2			CAI on *approximate* processing of numbers and quantities	TA		Hamburger Rechentest	0.45
Group 3			CAI on *exact and approximate* processing of numbers and quantities	TA		Hamburger Rechentest	0.21
Passolunghi & Costa (2016)	K	48	Counting, linear representation of numbers, relation between numbers and quantities, quantity comparison	TA		ENT	1.11
Ramani & Siegler (2008)	P	124	Number line board games, linear representation of numbers (The Great Race)	AR	Low-income; participants in Head Start	Number Line Estimation	0.67
Salminen, Koponen, Leskinen, Poikkeus, & Aro (2015)	K	21	CAI on counting, comparing, and simple addition	MD	Teacher referral of two lowest performing students per class; confirmed with performance >1.5 *SD*s below mean on basic fact screener	Number Sets Test	-0.46

Table 1 continued…

Study (year)	G[a]	*N*	Treatment	Group (risk)[b]	Criteria for Risk	Outcome Measure	Hedge's *g*
Schacter et al. (2016)	P	86	Subitizing, ordering quantities, counting, matching quantities	AR	Low-income; participants in Head Start	Number Sense	0.63
Schopman & Van Luit (1996)	K	60	Counting skills and early arithmetic skills	DIS	All students had documented disability; also required to score <45% correct on numeracy test	ENT	1.02
Siegler & Ramani (2009)	P	88	The Great Race	AR	Low-income; participants in Head Start	Number Line Estimation	0.63
Sood & Jitendra (2011)	K	101	Spatial relations, one and two more/less, benchmarks of 5 and 10, part-part-whole relationships (number sense instruction; NSI)				
Group 1				TA		Number Sense	1.38
Group 2			NSI	MD	<40th pc. on SESAT	Number Sense	1.13
Toll & Van Luit (2012)	K	192	Math language, reasoning, counting, numerals, number lines, simple calculations (The Road to Mathematics; TRTM)				
Group 1				MD	< 25th pc. ENT-R	ENT-R	0.26

Table 1 continued…

Study (year)	G[a]	*N*	Treatment	Group (risk)[b]	Criteria for Risk	Outcome Measure	Hedge's *g*
Group 2			TRTM	MD	25th–50th pc. ENT-R	ENT-R	0.90
Toll & Van Luit (2013)	P	276					
Group 1			TRTM	MD	Below average performance on ENT-R and typical working memory		1.29
Group 2			TRTM	MD-WM	Below average performance on ENT-R and <15th pc. on Automated Working Memory Assessment		1.08
Van de Rijt & Van Luit (1998)	K	106					
Group 1			Counting skills, solution strategies for basic arithmetic using *guiding* instruction; discovery learning (Additional Early Mathematics program; AEM)	MD	< 45% correct on Early Math Competence Scale	Early Math Competence Scales	1.25
Group 2			AEM with *structured* instruction; direct instruction	MD	< 45% correct on Early Math Competence Scale	Early Math Competence Scales	1.09

Table 1 continued…

Study (year)	G[a]	*N*	Treatment	Group (risk)[b]	Criteria for Risk	Outcome Measure	Hedge's *g*
Van Luit & Schopman (2000)	K	124	Conceptual counting	DIS	Documented disability and performance <25th pc. on ENT	ENT	0.74
Wilson, Dehaene, Dubois, & Fayol (2009)	K	53	CAI focused on numerical comparison, basic fact fluency	AR	Low-income	Symbolic Numerical Comparison	0.97

Note. Italicized *Groups* reflect that the groups are not independent, they share the same control group. pc. = percentile; ASPENS = *Assessing Student Proficiency in Early Number Sense* (Clarke, Rolfhus, Dimino, & Gersten, 2012); *Automated Working Memory Assessment* (Alloway, 2007); CBM battery = Curriculum-Based Measurement (CBM) Computation, *Addition Fact Fluency*, *Subtraction Fact Fluency*, and *CBM Concepts and Applications* (Fuchs, Hamlett, & Fuchs, 1990; Fuchs, Hamlett, & Powell, 2003); *Early Math Competence Scales* (derived from ENT); ENT = *Utrecht Test of Number Sense* (Early Numeracy Test; *ENT*; Van Luit, Van de Rijt, & Pennings, 1994); ENT (Finnish edition; Van Luit, Van de Rijt, Aunio, 2006; ENT-R = Van Luit & Van de Rijt (2009); *First grade concepts and applications* (Fuchs, Hamlett, & Fuchs, 1990); *Hamburger Rechentest* (Lorenz, 2007); *Missing Number* (MN; Clarke & Shinn, 2004); *Number Knowledge Test* (Okamoto & Case, 1996); *Number Sets Test* (Geary, Bailey, & Hoard, 2009); NSB = *Number Sense Brief* (Jordan, Glutting, Ramineni, & Watkins, 2010); *Quantity Discrimination* (QD; Clarke & Shinn, 2004); RAENS = *ROOTS Assessment of Early Numeracy Skills* (Doabler et al., 2012); SAT = *Stanford Achievement Test – Primary Level 1* (Gardner, Rudman, Karlsen, & Merwin, 1987); SESAT = *Stanford Early Achievement Test – Mathematics* (Madden, Gardner, & Collins, 1987); *TEDI-MATH* (Grégoire, Noel, & Van Nieuwenhoven, 2004); TEMA = *Test of Early Mathematics Ability* (PRO-ED, 2007); TEMA-3 = *Test of Early Mathematics Ability*, 3rd edition (Ginsburg & Baroody, 2003); TEMI-PM = *Texas Early Mathematics Inventories – Progress Monitoring* (University of Texas System & Texas Education Agency, 2007); *Word Problems* (Jordan & Hanich, 2000).

[a] G = Grade (P = preschool, K = kindergarten, 1 = first grade).

[b] Group; AR = at-risk, DIS = disability, MD = mathematics difficulty (author specified criteria), TA = typically achieving.

[c] Represents that the grade in the table (P, K) may not reflect grade stated in the manuscript due to re-coding of foreign studies.

Methodological characteristics. Across all treatment groups (*k* = 51), the average sample size was 113 participants (*SD* = 105). The average sample size for treatment was 49 participants (*SD* = 48; range = 8 to 207), and 44 (*SD* = 39; range = 7 to 206) for control. Authors reported attrition in 15 studies, and attrition ranged from 1.5% to 27.9% (*M* = 10.3%, *SD* = 7.8%). Regarding how effect sizes were calculated, most treatment effects (*k* = 46) were calculated as the "difference between the mean outcome for the intervention group and the mean outcome for the comparison group, divided by the pooled within-group standard deviation of the outcome measure" (WWC, 2014, p.22). The remaining treatment effects (*k* = 5) were calculated by dividing the difference between the scores for the treatment and control groups at post-test by the pooled standard deviation. Regarding assignment to treatment condition, the majority of studies (*n* = 27; 82%) used random assignment; in the remaining studies, studies used matching procedures (*n* = 4) or specified that non-random assignment was used (*n* = 1). One study did not report assignment procedures.

Regarding the nature of the treatment condition, many studies reported on treatment groups (*k* = 18) who received interventions that supplemented the core mathematics curriculum without replacing any part of the core curriculum and some studies reported that treatment groups (*k* = 11) received interventions that supplemented the regular mathematics curriculum by replacing part of regular mathematics instruction (e.g., the intervention took place during independent work time during regular mathematics instruction time). Fewer studies reported on treatment groups (*k* = 9) who received interventions that completely replaced (i.e., supplanted) the regular mathematics curriculum. For 13 treatment groups, it was not clear if the intervention supplemented or supplanted regular mathematics instruction.

Regarding the nature of the control condition, many studies had business-as-usual (BAU; i.e., regular mathematics instruction) control conditions ($k = 33$); however, some studies had BAU control conditions that controlled for intervention time (i.e., total time for intervention and control were equivalent; $k = 19$) and other studies had BAU control conditions that did not control for intervention time (i.e., the intervention group received supplemental support; $k = 14$). Twelve control conditions were active comparison groups that controlled for the time in which the control group received an alternative activity or intervention that was not related to mathematics (e.g., reading intervention), and two control conditions received an alternative mathematics intervention. Few studies did not report on the nature of the control condition ($k = 4$). The codes for nature of the treatment and control groups were the same for treatment groups across studies that included more than one treatment group.

Study quality. Five variables were considered for the quality of the study in this meta-analysis, including fidelity of implementation of the intervention, description of intervention agents, description of study participants with disaggregated information by group, description of the control condition (enough information to allow for replication), and attrition information. For each component of study quality, studies received a score of 0 or 1 (for a complete description of how studies were determined to meet criteria of receiving a 0 or 1, see Appendix A), and each study was given an overall total score that ranged from 0 (study did not address any study quality components) to 5 (study addressed all study quality components). The overall total score was the sum score of each of the five components.

The same study quality rating was applied to each treatment group in the study with studies with more than one treatment group; therefore, results are discussed in terms of studies in this section. Most studies reported treatment fidelity procedures for the intervention (n = 23), provided demographic information for intervention agents (n = 18), provided disaggregated participant demographics by treatment groups and control (n = 22), and described the control condition in a manner that allowed for replication (n = 18). About half of all studies reported attrition information (n = 16). When considering all study quality components, the average score on a scale of 0 to 5 for studies was 2.94 (SD = 1.43, range = 0 to 5). Four studies received a score of 5 (meaning they reported all information coded for study quality); eleven studies received a score of 4; seven studies received a score of 3; six studies received a score of 2 and, and six studies received a score of 1. Only one study received a score of 0, meaning that none of the information required for the study quality codes was reported.

Participant characteristics. Appendix C provides detailed information on the demographic information reported for each treatment group. Across all studies there was a total of 4,556 participants with 2,626 (58%) participants in treatment groups. Generally, studies reported demographic information for treatment and control groups separately; however, seven studies reported only full sample demographics. Most studies reported the average age of participants, gender, proportion of students receiving free or reduced price lunch (FRL), number of students identified as English Learners (ELs), and ethnicity representation, but five studies did not report demographics beyond the average age or grade level of participants. Approximately 51% of participants were male. Ethnicity was reported for approximately 73% of participants with representation of Caucasian (31%), African American (25%), Hispanic (13%), Asian (2%), and other (2%). With regard to U.S. studies only, approximately 47% of participants were identified as receiving FRL, and 22% identified as ELs.

Studies included participants who were in preschool (n = 7), kindergarten (n = 19), and first grade (n = 7). The average age of preschool participants was 4.74 years (*SD* = 0.48 years) for treatment and 4.74 years (*SD* = 0.48 years) for control; the average age of kindergarten participants was 5.54 years (*SD* = 0.44 years) for treatment and 5.53 years (*SD* = 0.36 years) for control; the average age of first-grade participants was 6.85 years (*SD* = 0.28 years) for treatment and 6.71 years (*SD* = 0.14 years) for control. Across all studies, the age of participants ranged from 4.4 years to 7.2 years.

Of the 51 treatment groups, approximately half (k = 27; 53%) met mathematics screening criteria to be considered at-risk for mathematics difficulty (MD) and an additional 6 treatment groups were considered at-risk for potential academic failure due to low-socioeconomic status (SES); meanwhile, 18 treatment groups were considered typically achieving. The most common method to identify students as at-risk for MD was with a specific cut score, percentile, or *SD* below the mean on a mathematics screening measure. Sixteen groups were identified with this method, and specific scores that authors reported for participants to be considered MD ranged from 1.5 *SD*s below the mean (performance below the 7th percentile, approximately) to performance below the 50th percentile. The most common percentile cutoff was performance below the 25th percentile. Five groups were identified as at-risk with progress monitoring measures that were administered over a period of time (i.e., the lowest performing students in a class or school, without a specific cutoff on a screening measure, were identified as students at-risk). Three groups were identified as at-risk for MD due to documented disabilities and receiving special education services; however, two of these groups also met percentile criteria on a screening measure. Finally, three groups were referred by classroom teachers to researchers as having MD for low performance; in two of these groups the teacher referrals were corroborated with follow-up screening on a mathematics measure. In addition to being identified as at-risk through the use of screening measures or teacher referrals, some participants were identified as having some academic risk due to low SES ($k = 6$). For example, in a kindergarten study, 91% of participants were eligible for FRL (Dyson et al., 2013), and in another preschool study all participants were considered at-risk due to meeting income requirements for entry to Head Start (Ramani & Siegler, 2008).

Intervention characteristics. Table 2 provides information specific to the intervention including numeracy domains addressed, instructional format, intervention agent, total min of instruction, nature of the treatment condition, and measurement of fidelity. On average, interventions were 11.5 weeks in duration ($SD = 9.5$ weeks), and an average of 28 sessions ($SD = 21$ sessions, range = 4 to 90 sessions). Intervention sessions lasted between 6 min and 60 min ($M = 26.5$ min, $SD = 10.6$ min), with an average total instruction time of 778 min ($SD = 645$ min). The frequency of the interventions ranged from 1 session per week to 6 sessions per week ($M = 3.0$ sessions; $SD = 1.2$ sessions).

With regard to the interventionist, trained researchers (e.g., graduate students, research assistants) delivered instruction in 14 studies (42%), followed by teachers or other school personnel in 12 studies (36%), and in 1 study the intervention agents were a mix of researchers and school personnel (3%). In five studies (15%), intervention content was delivered by computer-assisted instruction and one study did not report who delivered intervention instruction (3%). With regard to instructional grouping, the majority of studies (n = 20; 61%) used small-group instruction (i.e., groups of 2 to 5 students). Many studies (n = 8; 24%) also used one-to-one instruction; students either met independently with the intervention agent or they worked independently on a computer program. Of the remaining studies, three studies (9%) used peer-assisted learning strategies (PALS), one study (3%) used a mix of instructional formats (i.e., small group, whole class, and paired activities), and one study (3%) did not report the instructional format.

Table 2
Numeracy Domains and Intervention Characteristics

	Numeracy Domains			Intervention Characteristics				
Study (year)[a]	Number	Relations	Operations	Instructional format	Agent	Total min instruction	Nature of treatment	Fidelity measured
Bryant et al. (2011)	X	X	X	Small group	M	1900	Supplement	Yes
Clarke, Doabler, Smolkowski, Baker, et al. (2016)	X	X	X	Small group	SS	1000	Replaces part	Yes
Clarke, Doabler, Smolkowski, Kurtz Nelson, et al. (2016)	X	X	X	Small group	SS	1000	Supplement	Yes
Clarke et al. (2014)	X	X	X	Small group	SS	1800	Supplement	Yes
Clements (1984)	X	X	X	Small group	T	600 – 720	NR	No
Codding et al. (2011)	X	X	X	PA	R	480	Replaces part	Yes
Desoete & Praet (2013)								
Counting	X		X	Individual	CAI	225	NR	Yes
Comparison		X		Individual	CAI	225	NR	Yes
Doabler et al. (2016)	X	X	X	Small group	SS	1000	Supplement	Yes
Dyson et al. (2015)								
Fact retrieval	X	X	X	Small group	R	720	Supplement	Yes
Number sequence list	X	X	X	Small group	R	720	Supplement	Yes
Dyson et al. (2013)	X	X	X	Small group	R	720	Supplement	Yes

Table 2 continued…

	Numeracy Domains			Intervention Characteristics				
	Number	Relations	Operations	Instructional format	Agent	Total min instruction	Nature of treatment	Fidelity measured
Fuchs et al. (2005)	X	X	X	Small group	R	1920	Supplement	Yes
Fuchs et al. (2001)[a]	X	X	X	PA	T	600	Replaces part	Yes
Fuchs et al. (2002)[a]	X	X	X	PA	T	1440	Replaces part	Yes
Fuchs et al. (2013)								
Speeded practice	X	X	X	Individual	R	1440	Replacement	Yes
Non-speeded practice	X	X	X	Individual	R	1440	Replacement	Yes
Hansmann (2013)								
Traditional QD		X		Individual	R	125 – 175	NR	Yes
Triangle QD		X		Individual	R	125 – 175	NR	Yes
Hassinger-Das et al. (2015)								
Numeracy	X	X	X	Small group	R	720	Supplement	Yes
Story problems	X	X	X	Small group	R	720	Supplement	Yes
Jordan et al. (2012)	X	X	X	Small group	R	720	Supplement	Yes
Kaufmann et al. (2005)	X	X	X	Small group	T	900	Replacement	Yes
Kyttälä et al. (2015)	X	X		Small group	NR	240	NR	No
Mohanty & Mishra (1994)	X	X	X	Small group	T	600 – 750	NR	No

Table 2 continued…

	Numeracy Domains			Intervention Characteristics				
	Number	Relations	Operations	Instructional format	Agent	Total min instruction	Nature of treatment	Fidelity measured
Obersteiner et al. (2013)								
Exact		X	X	Individual	CAI	300	Supplement	No
Approximate		X	X	Individual	CAI	300	Supplement	No
Both		X	X	Individual	CAI	300	Supplement	No
Passolunghi & Costa (2016)	X	X		Small group	R	600	NR	Yes
Ramani & Siegler (2008)	X	X		Individual	R	60 – 80	NR	No
Salminen et al. (2015)	X	X	X	Individual	CAI	120 – 225	Supplement	Yes
Schacter et al. (2016)	X	X		Individual	CAI	180	NR	No
Schopman & Van Luit (1996)	X	X		Small group	NR	390	Replacement	No
Siegler & Ramani (2009) *Linear*								
Sood & Jitendra (2011)[a]	X	X	X	Flexible	T	400	Replaces part	Yes
Toll & Van Luit (2012)[a]	X	X	X	Small group	R	480	Replacement	Yes
Toll & Van Luit (2013)[a]	X	X	X	Small group	T	2700	Replacement	Yes

Table 2 continued…

	Numeracy Domains			Intervention Characteristics				
	Number	Relations	Operations	Instructional format	Agent	Total min instruction	Nature of treatment	Fidelity measured
Van de Rijt & Van Luit (1998)								
Guiding	X	X		Small group	R	780	Supplement	No
Structured	X	X		Small group	R	780	Supplement	No
Van Luit & Schopman (2000)	X	X	X	Small group	R	1440	Replacement	Yes
Wilson et al. (2009)	X	X	X	Individual	CAI	120	Supplement	No

Note. Flexible grouping = multiple methods of instructional grouping used. CAI = computer-assisted instruction; M = mix of researcher and teacher; PA = Peer-assisted; R = researcher, SS = school staff (other than teacher), T = teacher; NR = not reported; QD = quantity discrimination.

[a] Represents that there was more than one treatment group, but groups differed only on participant risk status; treatment was identical across groups.

Instructional features. Appendix D provides detailed information regarding the instructional features present in each intervention. Regarding explicit and systematic instruction characteristics, treatment groups received interventions that included the following components: scripted lesson plans (*k* = 36; 71%), corrective feedback (*k* = 28; 55%), modeling (*k* = 20; 39%), guided practice (*k* = 22; 43%), and independent practice (*k* = 30; 59%). Overall, 48 (94%) treatment groups received interventions that had at least one component of explicit and systematic instruction. Many treatment groups also received positive reinforcement during the intervention (*k* = 21; 41%); less often, treatment groups followed specific behavior management plans as part of the study (*k* = 4; 8%). Regarding the use of representations, 41 (80%) treatment groups received instruction that incorporated the use of visual representations and slightly fewer treatment groups (*k* = 34; 67%) received instruction that incorporated concrete representations. Twelve (24%) treatment groups received instruction that specifically used the concrete-representational-abstract (CRA) framework to teach numeracy content. While 15 (29%) treatment groups received instruction that *included* the use of games or books, 15 other treatment groups received interventions that were *based* solely on games or books.

Control condition. Studies were also coded with regard to features of the control condition, using the same codes as those for coding the treatment condition. Due to the dependency of some treatment groups on the same control group, there were only 43 control groups to consider when coding the studies. Many studies did not describe the control condition in a manner that would allow for replication of the control setting (*k* = 25; 58%). This included instances where no information was given about the instructional features of the control condition and instances where studies simply reported the control condition as free play or center time without any further description of those conditions or settings.

Control groups were described with some detail to allow for replication of the condition in 18 studies (42%). These control groups received instruction that included content from the Number domain (*k* = 17; 40%), the Relations domain (*k* = 12; 28%), and the Operations domain (*k* = 7; 16%). The most common skills addressed in the control condition included counting sequence, counting with one-to-one correspondence, and numeral identification. Some studies (*k* = 14; 33%) also provided details regarding instructional features. The most common instructional features present in the control setting included concrete representations, worksheets, and explicit and systematic instruction.

Dependent measures. Appendix E provides information regarding the primary mathematics measures used to calculate the effect size for this meta-analysis. All studies included at least one mathematics outcome measure. Of the 51 treatment groups, 21 (41%) were assessed with only one mathematics measure to determine treatment effects. The remaining groups were administered more than one mathematics measure to determine the effectiveness of the numeracy intervention. Regarding the primary measure that was selected for each study, 27 (53%) treatment groups used a broad proximal measure, 15 (29%) treatment groups used a broad distal measure, 5 (10%) treatment groups used a narrow proximal measure, and 4 (8%) treatment groups used a narrow distal measure. More than half of the treatment groups (*n* = 28; 55%) were assessed using a researcher-developed measure (that was not norm-referenced) as the primary outcome measure to determine treatment effects. Slightly fewer treatment groups (*n* = 20; 39%)

were assessed using a norm-referenced measure. This result was not surprising given that primary measures that were more closely aligned to the intervention (typically researcher-developed) were selected over distal measures of mathematics performance, such as norm-referenced tests of mathematics achievement. Three treatment groups (6%) that were in the same study were assessed with a measure that was not clearly identified as either norm-referenced or researcher-developed.

Information for reliability and validity of primary measures was also recorded. Of the 51 primary measures used, authors reported reliability estimates for 39 measures (71%). In most cases, reliability estimates were reported in the form of Cronbach's alpha. Significantly fewer authors reported information regarding the validity estimates for measures. Authors reported validity estimates for only 12 primary measures (22%), possibly reflecting the large number of researcher-developed measures used that may have lacked validity information.

Regarding the total number of measures and type of measure given to each treatment group, on average, treatment groups were assessed with approximately three mathematics measures (range = 1 to 7 measures). Most treatment groups ($k = 33$; 65%) were not administered non-mathematics measures. Of the remaining studies where treatment groups were administered non-mathematics measures ($k = 18$; 35%), the number ranged from 1 to 10 non-mathematics measures. Examples of the types of non-mathematics measures administered to treatment groups included measures of attention, visual and verbal working memory, letter identification, and listening comprehension.

Confidence codes. Following all of the coding for study information, methodological characteristics, participant characteristics, intervention characteristics, and dependent measures, studies were coded with regard to confidence in those codes. General confidence codes were given to each treatment group, not necessarily to each study, as some discrepancies were specific to information provided about a certain treatment group. The majority of treatment groups received a code of 0 (k = 31; 61%), indicating zero or one perceived discrepancy in the information provided in the manuscript. The remaining treatment groups (k = 20; 39%) received a code of 1, meaning there were two to four perceived discrepancies in the information provided in the manuscript. No treatment groups received a code of 2, as no studies had five or more discrepancies in information reported.

Confidence codes were also assigned to each study regarding the confidence in the estimate of the effect size. The vast majority of treatment groups (k = 48; 94%) provided conventional statistics (means, *SD*s, sample size) to calculate effect sizes. These studies received a score of 0 for no estimation. The remaining 3 treatment groups received a code of 1, for some estimation, as the results that were reported required conversion from *t*-test values to effect sizes. No treatment groups received a score of 2, as no studies were considered to have high estimation of effect sizes.

Meta-analysis Results

In the next section, I explain the process for identifying potential outliers. Results are discussed for each research question with and without potential outliers. Effect sizes are interpreted as small (g = 0.20), moderate (g = 0.50), and large (g = 0.80; Cohen, 1988). This section concludes with results of the publication bias analyses.

Addressing potential outliers. The distribution of effect sizes for the total sample is provided in Figure 1; the distribution of effect sizes for the sample with the two outliers removed is provided in Figure 2. The distribution of effect sizes appeared normal after removing the two outliers. I identified potential outliers as those effect sizes that were more than 3.0 *SDs* above the random weight mean effect size (i.e., effect sizes larger than $g = 2.11$ for this sample of studies; Cooper et al., 2009). Two studies were identified as potential outliers with effect sizes estimated at $g = 4.31$ (Clements, 1984) and $g = 6.63$ (Mohanty & Mishra, 1994). The results for the research questions below are first discussed including the potential outliers, followed by results excluding the potential outliers.

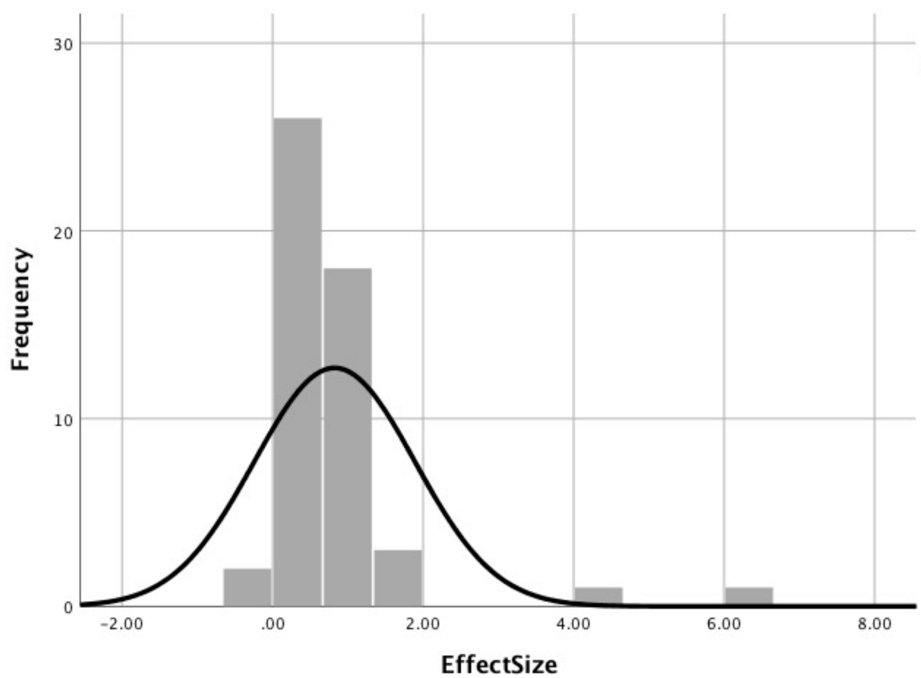

Figure 1. Distribution of effect sizes (total sample)

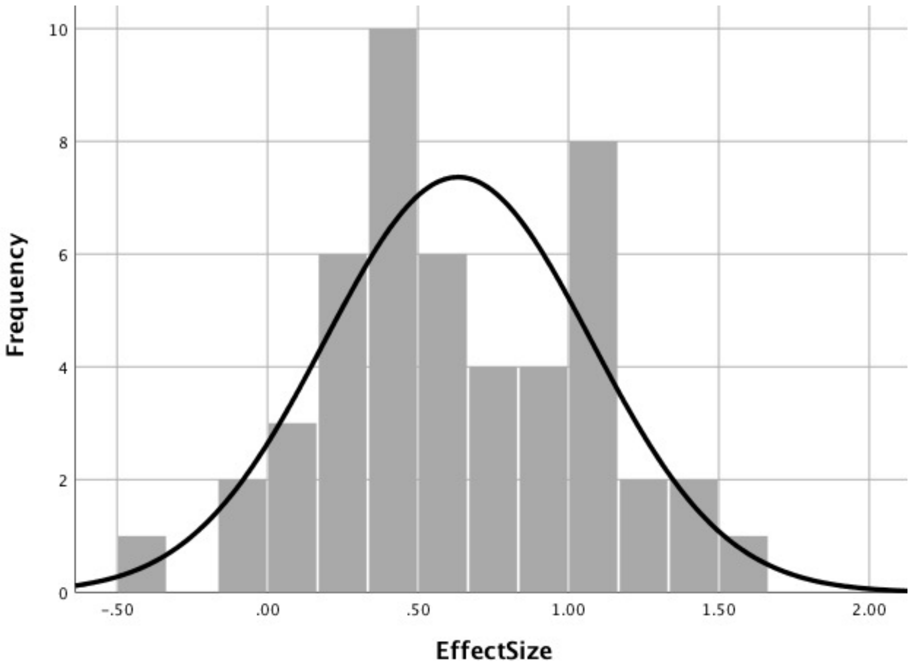

Figure 2. Distribution of effect sizes (outliers removed)

Research question 1. *What is the overall mean effect, and how variable are those effects, of early numeracy interventions on mathematics outcomes for students in preschool, kindergarten, and first grade on proximal outcome measures? On distal outcome measures?*

The forest plot in Figure 3 below shows the effect size estimates and 95% confidence interval for the total sample; Figure 4 is the forest plot with the two outliers removed.

Weighted random mean effect size. The overall weighted random mean effect size of 0.68 (*SE* = 0.07; range = -0.46 to 6.63) for all early numeracy interventions in this meta-analysis was moderate-to-large, and the 95% confidence interval did not include zero [0.54, 0.81]. Nineteen groups yielded large effects, five groups reported moderate-to-large effects, 11 groups reported moderate effects, 13 groups reported small effects, and 3 studies reported negative effects for treatment groups (see Table 1 for effect size for each treatment group). The test of heterogeneity was significant ($Q_{(50)} = 220.38$; $p < .001$) and approximately 77% of the variability in effect sizes was due to heterogeneity rather than sampling error (I^2). After removing the two potential outliers, the overall weighted mean effect size was slightly lower and moderate at 0.63 ($SE = 0.06$), with a 95% confidence interval that did not include zero [0.50, 0.73]. The test of heterogeneity remained significant ($Q_{(48)} = 159.77$; $p < .001$; $I^2 = 70\%$).

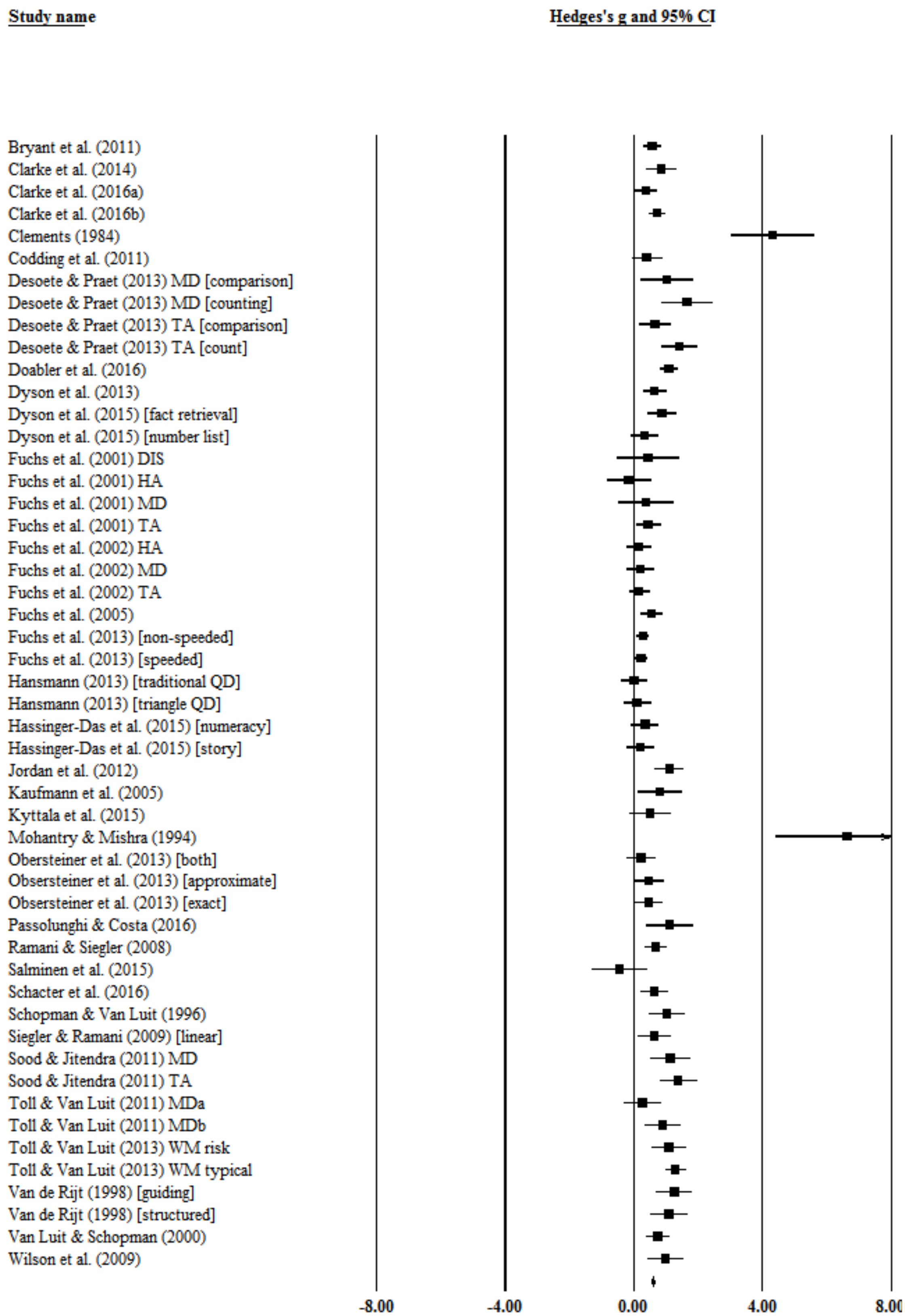

Figure 3. Forest plot (total sample)

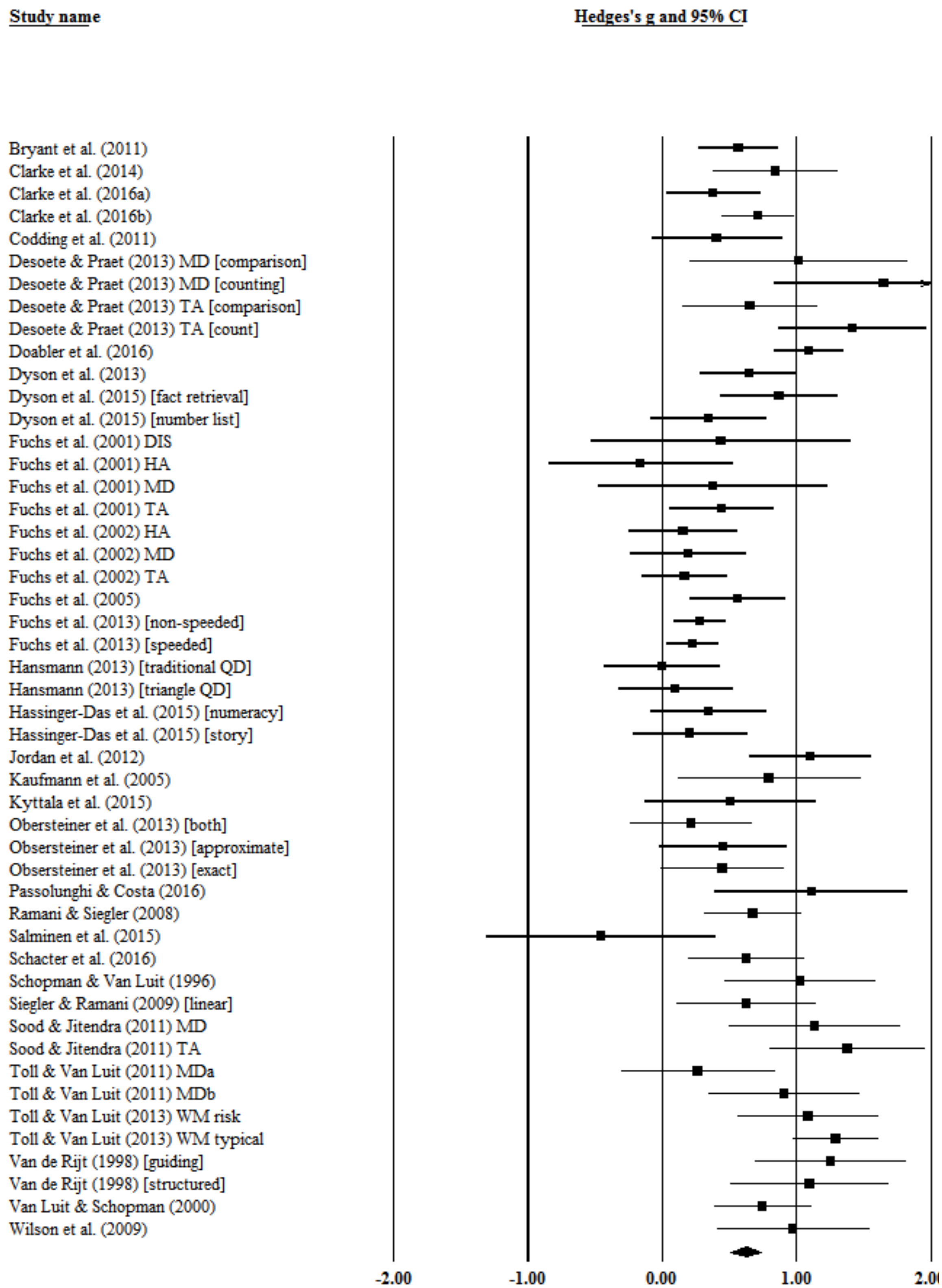

Figure 4. Forest plot (outliers removed)

Sensitivity analysis. One treatment group from each of the dependent pairs of treatment groups was selected as the treatment group that was more closely aligned to the early numeracy skills that were coded for this meta-analysis (i.e., the "selected" treatment group; see Appendix B for information regarding which treatment group was selected). Table 3 summarizes the results of the sensitivity analysis both with and without the two potential outliers. The results indicate that "selected" and "non-selected" treatment groups yielded very similar results when considering those sets of treatment groups combined with all of the independent effects; therefore, all treatment groups, including those with dependent effects, were included in this meta-analyses.

Table 3
Summary of Sensitivity Analysis Results

				95% Confidence Interval	
Comparison	*k*	*g*	*SE*	Lower	Upper
Full Sample (*k* = 51)					
Independent groups and selected groups	43	0.72	0.08	0.57	0.88
Independent groups and non-selected groups	43	0.71	0.08	0.56	0.86
Selected groups only	7	0.51	0.15	0.21	0.80
Non-selected groups only	7	0.46	0.15	0.16	0.75
Outliers Removed (*k* = 49)					
Independent groups and selected groups	41	0.66	0.07	0.53	0.79
Independent groups and non-selected groups	41	0.65	0.06	0.53	0.77

Note. The analysis with "Selected groups only" and "Non-selected groups only" was not repeated for the Outliers Removed because none of the dependent treatment groups were outliers.

Dependent measure. Appendix F provides the mean effect sizes, standard error, 95% confidence interval, and the within-group *Q* test results for the outcome measures. In addition to the type of outcome measure (broad, narrow, proximal, distal), information regarding whether the primary outcome measure was norm-referenced is also presented.

Not all studies included both a proximal outcome measure and a distal outcome measure. In fact, 21 treatment groups were assessed with only one mathematics outcome measure, so comparing overall results (not just the primary measure results) of treatment effects on proximal and distal measures for each study was not appropriate in this meta-analysis. Thus, results for the treatment effect contingent on the type of outcome measure are discussed in terms of the primary measure only. Although treatment groups (*n* = 4; 8%) that were administered distal narrow measures of mathematics with regard to the measure's alignment to the intervention produced the largest effect (*g* = 1.16; *SE* = 0.24), this set of treatment groups were from the same study. Not including this study, treatment groups who were administered broad proximal measures produced the largest effect (*g* = 0.84; *SE* = 0.08). Treatment effects also differed according to whether or not treatment groups were assessed with a norm-referenced measure; treatment groups (*k* = 28; 55%) who were assessed with researcher-developed measures produced a larger average effect (*g* = 0.84; *SE* = 0.09) compared to treatment groups (*k* = 20; 39%) who were assessed with norm-referenced measures (*g* = 0.50; *SE* = 0.11). Results between the full sample (*k* = 51) and sample with outliers removed (*k* = 49) were similar with regard to outcome measures (see Appendix F).

Research question 2. *Which early numeracy domain was most investigated, and which domain produced the largest effect size?*

Frequency of numeracy domains. As set by the inclusion criteria, all interventions included a skill from at least one early numeracy domain (Number, Relations, and Operations); Table 2 above provides information regarding which numeracy domains were included in each treatment group intervention. The most investigated early numeracy domain was the Relations domain (*k* = 49; 96%); slightly fewer treatment groups received interventions that included content that addressed the Number domain (*k* = 44; 86%). The least investigated early numeracy domain was the Operations domain, though many treatment groups still received interventions with this content (*k* = 39; 76%). In addition to specific early numeracy domains addressed, 10 (20%) treatment groups received interventions that explicitly included a mathematics language component (e.g., intentionally included mathematics vocabulary as part of the intervention). Figure 5 below shows the representation of numeracy skills and mathematics language by grade for each treatment group, keeping in mind that figure depicts frequency of skills in grade-levels that did not have equal sample sizes of treatment groups (preschool, *n* = 8; kindergarten, *n* = 30; first grade, *n* = 13).

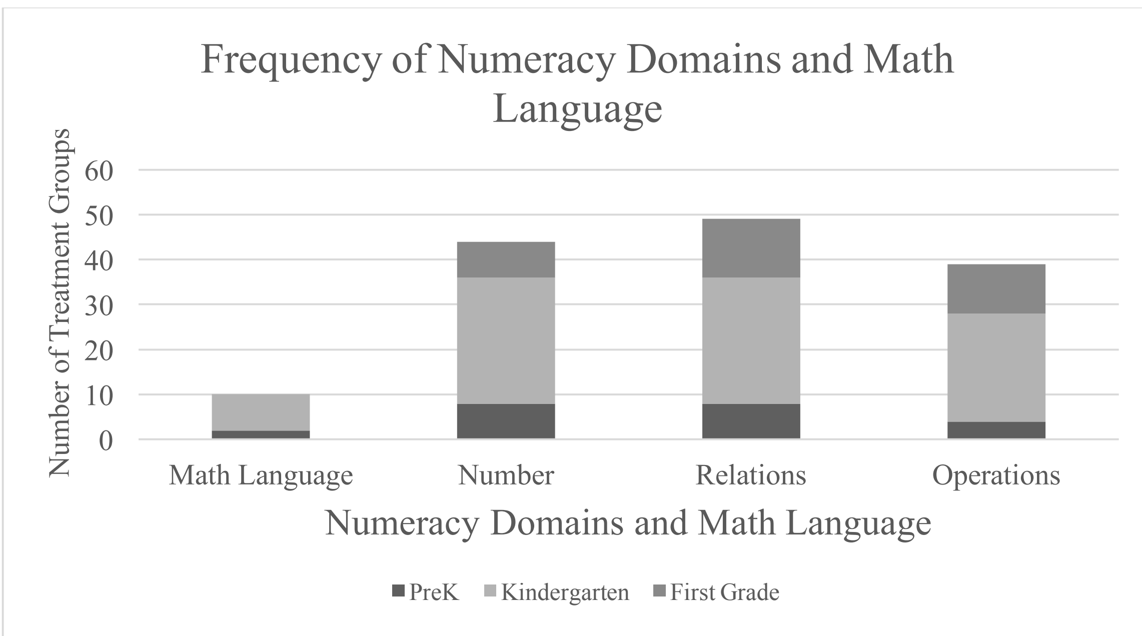

Figure 5. Bar graph illustrating the frequency of numeracy domains and mathematics language.

Effect of numeracy domains. To determine which domain produced the largest effect, weighted average effect sizes were calculated based on if a study included content in a specific domain. In addition, because there was little variation in coding treatment groups dichotomously (yes, no) for receiving intervention content in a specific numeracy domain (i.e., most interventions included at least one skill in the Number and Relations domain), the number of different skills addressed in each intervention was also explored to provide more insight regarding features of effective early numeracy interventions. Those results are also presented here.

Appendix G provides the mean effect sizes, standard error, 95% confidence interval, and the within-group *Q* test results related to the treatment effects for numeracy domain representation and number of skills within each domain. Results are presented for the full sample (k = 51) and the sample with outliers removed (k = 49). Interventions with Number domain content had the largest weighted mean effect size (g = 0.72; *SE* = 0.07), followed by interventions with Operations domain content (g = 0.67; *SE* = 0.08) and Relations domain content (g = 0.65; *SE* = 0.07). Each domain yielded a moderate-to-large treatment effect, and results between the full sample and sample with outliers removed were similar with each domain producing moderate effects. Interventions that specifically included a mathematics language component yielded a large effect (g = 0.81; *SE* = 0.15), with no difference after outlier removal.

Number of skills in each numeracy domain. When exploring number of skills addressed in each intervention further, there was more variability in those results compared to the results of the dichotomous coding (also presented in Appendix G). In the Number domain, interventions that addressed two or three skills had the largest average effect (*g* = 0.97; *SE* = 0.10); the 95% confidence interval did not have any overlap with interventions that addressed zero or one Number skills and had very little overlap with interventions that addressed four or more Number skills. Interventions that addressed zero or one skills or four or more skills had small and moderate average effects, respectively. A different pattern emerged for interventions that addressed skills in the Relations domain. Unlike interventions in the Number domain, interventions that addressed two or three Relations skills had the smallest average effect (*g* = 0.59; *SE* = 0.09); however, this average effect was still moderate. Interventions in the Relations domain that addressed few (zero, one) and many (four or more) skills had similar, large average effect sizes (*g* = 0.80; *SE* = 0.15 and *g* = 0.78; *SE* = 0.14, respectively) and overlapping 95% confidence intervals. Finally, interventions that addressed skills in the Operations had the largest effect when interventions addressed zero or one skills (*g* = 0.84; *SE* = 0.10), followed by interventions that addressed four or more skills (*g* = 0.60; *SE* = 0.14). The smallest effect for the Operations domain was observed for interventions that addressed two or three skills (*g* = 0.49; *SE* = 0.12).

Frequency of skills in numeracy domains. To provide a more in-depth evaluation of the numeracy domains, individual skills within each domain were also examined with regard to the frequency with which they were included in interventions and the average effect of interventions that included specific skills. Appendix H provides detailed information regarding the specific numeracy skills addressed by each study. Across the three domains, the seven most common skills addressed by intervention lessons (each skill was included in more than half of interventions) included: counting sequence (76%), number line sequences (69%), one-to-one counting (69%), numeral comparison (60%), numeral identification (60%), simple addition and subtraction with objects (58%), and set comparison (56%). Figures 6, 7, and 8 below show the frequency of early numeracy skills by domain and grade of participants.

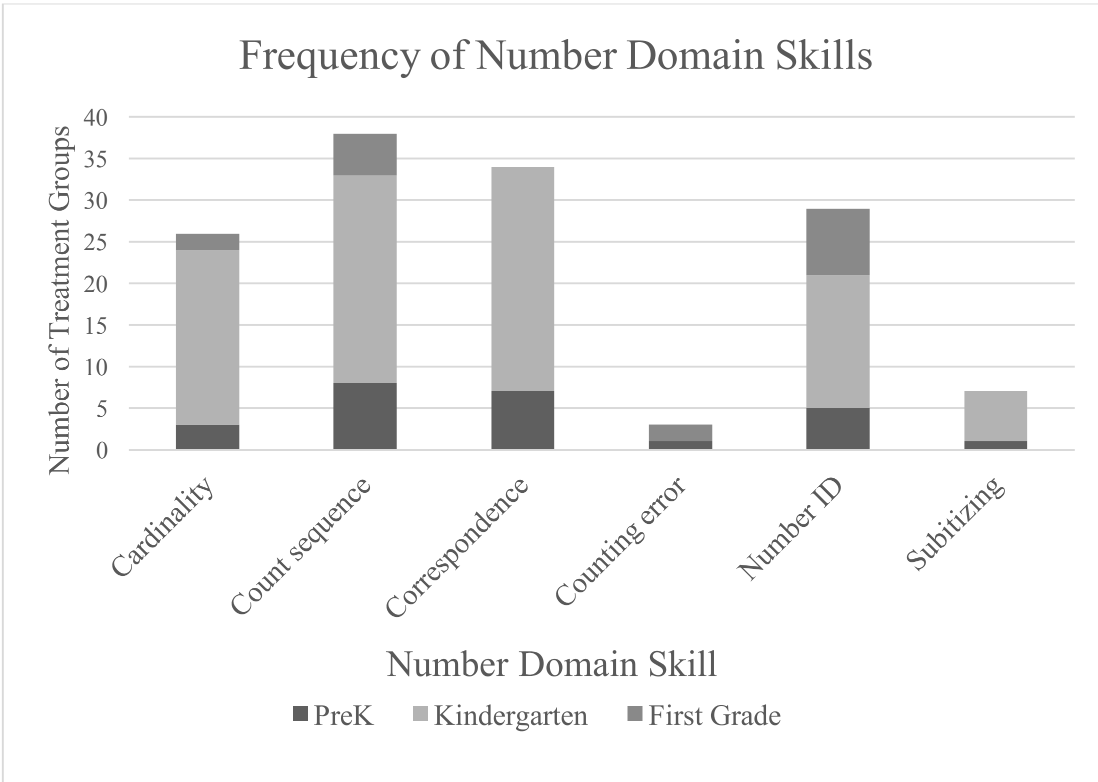

Figure 6. Frequency of Number domain skills

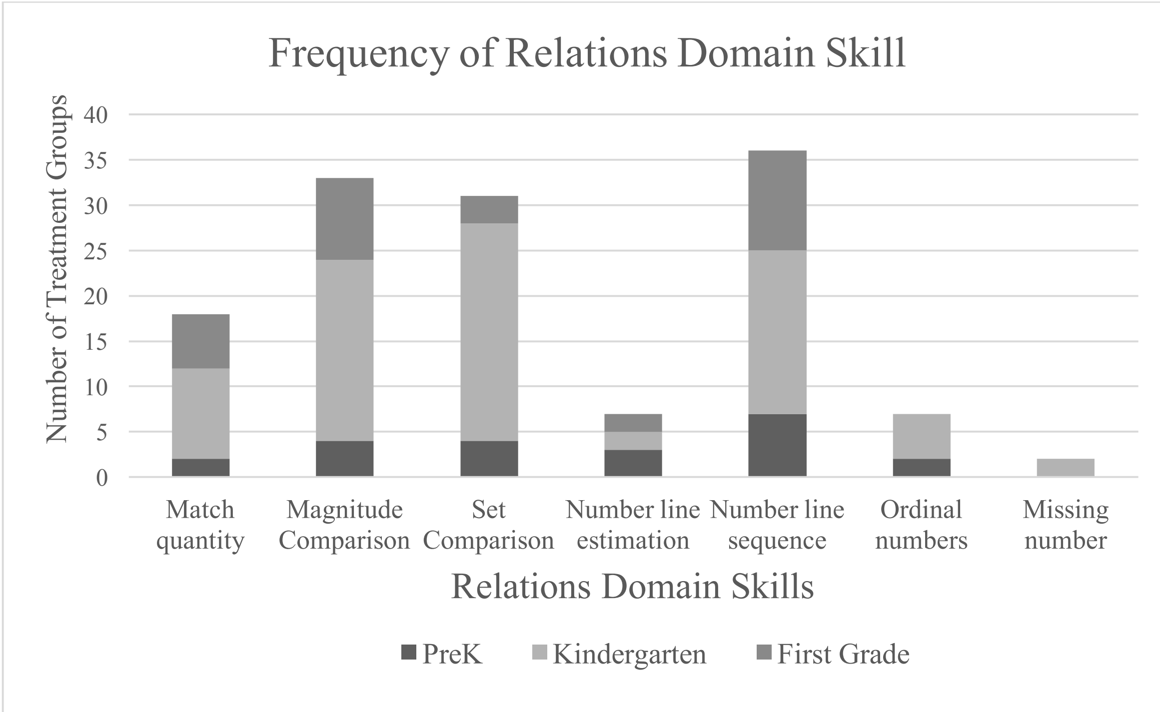

Figure 7. Frequency of Relations domain skills

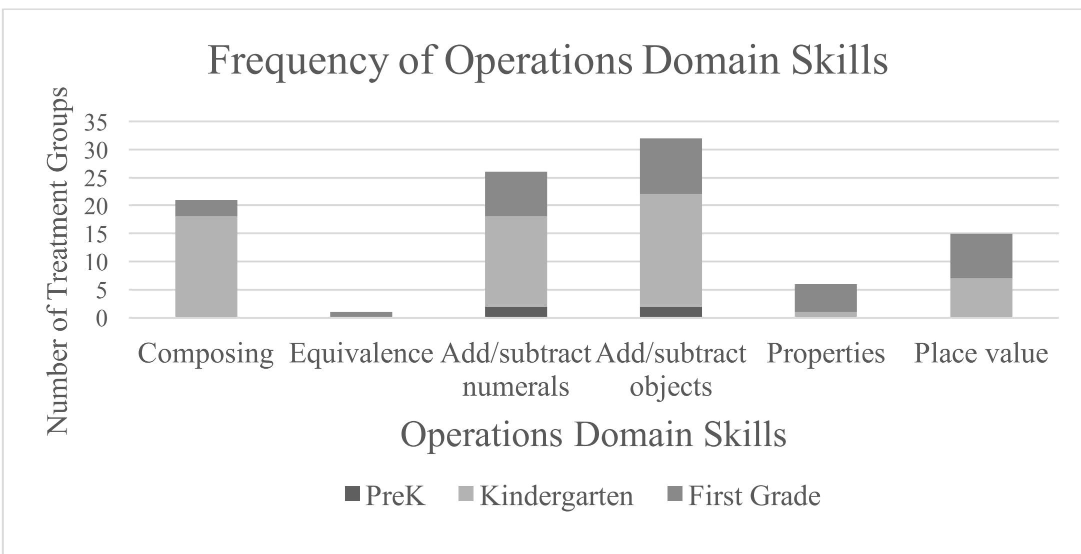

Figure 8. Frequency of Operations domain skills

Effect of specific skills in numeracy domains. Appendix I provides the mean effect sizes, standard error, 95% confidence interval, and the within-group *Q* test results for skills addressed in the Number, Relations, and Operations domains. With regard to interventions with skills in the Number domain, all early numeracy skills yielded moderate-to-large effect sizes ranging from 0.57 to 0.85. Interventions that included counting with one-to-one correspondence (g = 0.85; SE = 0.07) and cardinality (g = 0.83; SE = 0.01) reported the largest mean effect sizes. Outlier removal yielded similar results for all skills with the exception of counting error skill. An outlier was one of only three studies to report including this skill in the intervention, so its removal produced a small average effect (g = 0.25; SE = 0.23). Skills in the Relations domain produced moderate-to-large effects, ranging from 0.59 to 0.97, and interventions that included ordinal number skills produced the largest effect (g = 0.97; SE = 0.17). Outlier removal produced similar results. Finally, skills in the Operations domain produced moderate effects, ranging from 0.51 to 0.65, and interventions that included skills in simple addition and subtraction with objects yielded the largest effect (g = 0.65; SE = 0.09). Outlier removal produced similar results.

Across all domains and skills, only three skills yielded average effects that included zero in the 95% confidence interval. Each skill had fewer than three interventions associated with it. These skills included counting error in the Number domain 95% CI [-0.19, 0.69], missing number in the Relations domain 95% CI [-0.10, 1.29], and equivalence in the Operations domain 95% CI [-0.32, 1.44]. Outlier removal yielded similar results.

Research question 3. *What are the differential treatment effects of early numeracy interventions on mathematics outcomes across study characteristics, participant characteristics, and intervention characteristics?*

To examine the potential causes in variation of the overall mean effect, the weighted average effect sizes and 95% confidence intervals for subcategories (e.g., participants with MD versus participants without MD) were calculated for specific variables that were coded. Then, between-groups variation was examined (i.e., *Q* test of homogeneity). A significant between-groups *Q* test result suggests that the variable may potentially explain variation in the overall mean effect. For clarity, only significant between-groups differences are discussed in-depth in text; however, all results are reported in the appendices associated with each section.

Study information. Appendix J provides the mean effect sizes, standard error, 95% confidence interval, and the within-group and between-groups *Q* test results for study information, including year of publication and location. Location yielded significant between-groups variation. Results for the full sample reported that studies published in the U.S. produced a smaller average effect ($g = 0.55$; $SE = 0.08$) compared to studies published outside the U.S ($g = 0.88$; $SE = 0.11$). Results of the between-groups *Q* test were significant ($p(Q) = 0.01$) and remained significant after removing outliers ($p(Q) = 0.01$). The between-groups *Q* test was not significant for date of publication when comparing studies that were published in the last 5 years and studies that were published in 2011 or earlier ($p(Q) = 0.34$); removing outliers also produced non-significant results.

Methodological characteristics. Appendix K provides the mean effect sizes, standard error, 95% confidence interval, and the within-group and between-groups *Q* test results for methodological characteristics, including independence of treatment effects, effect size calculation, assignment to the treatment condition, nature of the treatment and control, study quality, and confidence codes. Results for methodological characteristics indicate that none of the variables examined in this meta-analysis produced significant between-groups *Q* test results when considering the full sample of studies and the sample with outliers removed.

Participant characteristics. Appendix L provides the mean effect sizes, standard error, 95% confidence interval, and the within-group and between-groups *Q* test results for participant characteristics including risk status and grade. Other participant characteristics (e.g., race, gender) were not examined due to missing data. Results for the grade level of participants reported significant between-groups variation. Preschool interventions reported the largest mean effect (g = 1.10; SE = 0.16), followed by kindergarten interventions (g = 0.75; SE = 0.08), and first grade interventions (g = 0.32; SE = 0.11). The between-groups *Q* test was significant ($p(Q) < .001$) and remained significant after removing the two outliers ($p(Q) < .001$). The between-groups *Q* test was not significant regarding the treatment effects for participants based on risk status ($p(Q)$ = 0.52); results approached significance after outliers were removed ($p(Q)$ = 0.12).

Intervention characteristics. Appendix M provides the mean effect sizes, standard error, 95% confidence interval, and the within-group and between-groups *Q* test results for intervention characteristics including duration (i.e., weeks of intervention), instructional arrangement, and intervention agent. Regarding the full sample, results indicate that only instructional arrangement yielded significant between-groups variation ($p(Q) < 0.001$). Most studies included treatment groups who received instruction in small groups (k = 25); on average, these treatment comparisons yielded large treatment effects (g = 0.84; SE = 0.09). Many treatment groups also received instruction in a one-to-one setting ($k = 14$) and these treatment groups produced moderate effects ($g = 0.60$; $SE = 0.11$). A few studies included treatment groups ($k = 8$) who received instruction using peer-assisted learning strategies (PALS); these treatment groups produced small effects on average ($g = 0.25$; $SE = 0.16$). Two treatment groups utilized instructional arrangements that were flexible groups (i.e., mix of instructional arrangements) and two treatment groups did not report instructional arrangements. On average, studies with treatment groups that utilized flexible grouping yielded large effects ($g = 1.28$; $SE = 0.33$) and studies with treatment groups that did not report instructional arrangement information yielded small effects ($g = 0.05$; $SE = 0.29$). Results for instructional arrangement were substantially similar or the same after removing two outliers for one-to-one instruction, flexible groups, PALS, and not reported. Two outliers were removed from studies with treatment groups that utilized small group instruction and similar results were also reported ($g = 0.76$; $SE = 0.07$). Finally, the between-groups Q test remained significant after removing two outliers ($p(Q) < 0.001$).

Instructional features. Appendix N provides the mean effect sizes, standard error, 95% confidence interval, and the within-group and between-groups Q test results for instructional characteristics including features of explicit and systematic instruction, representations, and if the intervention was game or book-based. Regarding the full sample, results indicated that several variables yielded significant between-groups variation including: the intervention had at least one component of explicit and systematic instruction ($p(Q) < 0.05$), scripted lessons ($p(Q) < 0.05$), pictorial representations ($p(Q) < 0.05$), and included books or games ($p(Q) < 0.10$). After outliers were removed, only pictorial representations ($p(Q) = 0.10$) and included books or games

($p(Q) = 0.10$) remained significant; however, larger average effect sizes were observed for the treatment groups who were not exposed to pictorial representations ($g = 0.84$; $SE = 0.15$) and were not exposed to books or games ($g = 0.75$; $SE = 0.08$) during the intervention, compared to treatment groups who did receive interventions that included pictorial representations ($g = 0.58$; $SE = 0.06$) and games or books ($g = 0.51$; $SE = 0.11$).

Finally, after removing outliers, treatment groups who received interventions using a CRA framework yielded larger average effects ($g = 0.81$; $SE = 0.12$) than groups who did not receive instruction in this framework ($g = 0.57$; $SE = 0.07$); this between-groups variation was significant ($p(Q) < 0.10$).

Specific numeracy skills. The results of the previous research question addressed the mean effect sizes, standard error, 95% confidence interval, and the within-group *Q* test results for skills addressed in the Number, Relations, and Operations domains; therefore, only between-groups variation is discussed here. Appendix I provides the between-groups *Q* tests results. Prior to removing two outliers, skills in the Number domain that yielded significant ($p < .10$) between-groups variation included cardinality, correspondence, counting sequence, and numeral identification. After removing outliers, correspondence was the only Number domain skill to maintain significant between-groups variation, with counting sequence approaching significance. Counting error had significant between-groups variation after outlier removal; however, only two treatment groups received instruction that included this skill.

In the Relations domain, number line sequence, ordinal numbers, and quantity discrimination skills yielded significant ($p < .10$) between-groups variation prior to outlier removal. After removing outliers, number line sequence and ordinal numbers skills remained significant. Interestingly, although significant in variation between-groups, interventions that included number line sequence ($g = 0.56$; $SE = 0.07$) yielded a smaller effect than interventions that did *not* include the skill ($g = 0.79$; $SE = 0.11$); however, including number line sequence skills still produced a moderate effect.

In the Operations domain, prior to outlier removal the only skill to yield significant between-groups variation was place value. After outlier removal, no skills in the Operations domain yielded significant between-groups variation.

Dependent measures. The results of research question 1 addressed the mean effect sizes, standard error, 95% confidence interval, and the within-group *Q* test results for outcome measures; therefore, only between-groups variation is discussed here. Appendix F also provides the between-groups *Q* test results for dependent measure characteristics, including the type of primary outcome measure (proximal or distal and broad or narrow) and whether the primary outcome measure was norm-referenced. Regarding the full sample and sample with outliers removed, results indicate that different primary outcome measure types had significant between studies variation ($p(Q) < 0.001$). Results were also significant regarding whether or not the primary measure was norm-referenced in the full sample of treatment groups ($p(Q) < 0.05$).

Research question 4. *Which variables accounted for the most between-studies variance for the total sample? For the three domains separately?*

As previously discussed, I determined that two studies (Clements, 1984; Mohanty & Mishra, 1994) were outliers; therefore, they were not included in the sample of studies used to conduct the meta-regression analyses. The purpose of this research question was to determine which set of variables would represent the final model and together accounted for between-studies variance for the total sample of studies, and for the three domains separately. To select the final model, several aspects were considered, including (a) practical significance of the variable (i.e., was the variable a malleable factor that teachers could easily control), (b) the results of previous research that identified variables as potentially influencing academic achievement, and (c) the results of the first three research questions that indicated variables may have some influence on the average weighted effect size for this sample of studies. First, the results for the full sample with outliers removed is discussed. Then, the exploration of the three early numeracy domain meta-regression results is discussed.

Variance accounted for in total sample. Table 4 provides a summary of the final model, including the predictors CRA, intervention duration, risk status of participants, and the inclusion of counting with one-to-one correspondence in the intervention content. CRA and counting with one-to-one correspondence were dichotomous variables coded as yes (1) or no (0). Intervention duration was coded as 8 weeks or less of total instruction (0) and more than 8 weeks of instruction (1). Risk status of participants was coded as low academic risk (0), risk of MD based on performance criteria that did *not* include performance below 25th percentile (1; e.g., performance at the 40^{th} percentile, percent correct on a screener that did not align to percentiles, students with disabilities who were not screened with a mathematics measure), risk of MD based on performance at or below the 25th percentile on a mathematics screening measure (2), and some academic risk based on low SES of participants (3).

Table 4
Summary of the Final Weighted Least Squares Meta-Regression Model

			95% Confidence Int.			
Covariate	*B*	*SE*	Lower	Upper	Z-value	*p*-value
Intercept	0.34	0.09	0.16	0.53	3.71	< 0.01
CRA	0.18	0.11	-0.03	0.38	1.67	0.08
Duration	-0.24	0.11	-0.46	-0.02	-2.18	0.03
Correspondence	0.51	0.10	0.33	0.70	5.40	< 0.01
MD Risk1 (>25th)	0.28	0.12	0.04	0.51	2.29	0.02
MD Risk2 (≤25th)	-0.24	0.13	-0.49	0.02	-1.81	0.07
Some Risk (SES)	-0.10	0.16	-0.41	0.22	-0.61	0.54
Final Model						
Tau^2	0.027					
I^2	35.95%					

Note. *K* = 49. For MD Risk1, MD Risk2, and Some Risk $Q_{(3)} = 13.72$, $p < 0.01$).
Based on the final model, the predicted treatment effect is represented as $T_{(treatment\ effect)} = 0.34 + 0.18(CRA) - 0.24(Duration) + 0.51(Correspondence) + 0.28(MD\ Risk1) - 0.24(MD\ Risk2) - 0.10(Some\ Risk)$
CRA = concrete-representational abstract framework; MD = mathematics difficulty; SES = socio-economic status; Correspondence = counting with one-to-one correspondence; Duration = weeks of intervention.

The coefficient for *Duration* was negative, indicating that interventions that were longer than 8 weeks yielded smaller treatment effects. The *MD Risk2* and *Some Risk* covariates were also negative, while the *MD Risk1* covariate was positive. With the *Low-risk* group serving as the reference group (i.e., students identified as typically achieving or high achieving), this result indicates that treatment effects were smaller for (a) students identified as at-risk according to low SES and (b) students at a greater degree of risk for MD according to performance on a mathematics screener (i.e., students at or below the 25th percentile are at a greater risk for MD than students with performance above the 25th percentile). In contrast, studies that included students with some degree of risk (*MD Risk1*) yielded larger treatment effects than students with *low-risk*. The coefficients for *CRA* and *Correspondence* were positive, indicating that interventions that included these intervention components yielded larger treatment effects on average. The model was statistically significant (Q_M (6) = 48.15, $p < 0.001$; Pseudo R^2 = 75%).

Sensitivity analysis. Table 5 summarizes the results of the sensitivity analysis for the final model. Selected treatment groups were determined to be more closely aligned to the early numeracy skills of interest in this meta-analysis compared to the non-selected treatment groups.

Table 5
Summary of the Sensitivity Analysis for Dependent Effects

Selected Treatment Groups			95% Confidence Int.			
Covariate	*B*	*SE*	Lower	Upper	Z-value	*p*-value
Intercept	0.24	0.13	-0.01	0.48	1.88	0.06
CRA	0.21	0.11	0.00	0.42	1.90	0.06
Duration	-0.15	0.12	-0.39	0.09	-1.19	0.23
Correspondence	0.57	0.11	0.36	0.79	5.28	< 0.01
MD Risk1 (>25th)	0.27	0.13	0.02	0.51	2.09	0.04
MD Risk2 (≤25th)	-0.36	0.15	-0.66	-0.05	-2.32	0.02
Some Risk (SES)	-0.05	0.17	-0.38	0.28	-0.30	0.76
Tau^2	0.026					
I^2	33.21%					
Non-selected Treatment Groups						
Intercept	0.36	0.11	0.14	0.59	3.19	< 0.01
CRA	0.16	0.11	-0.04	0.37	1.53	0.12
Duration	-0.24	0.12	-0.48	-0.01	-2.00	0.05
Correspondence	0.46	0.10	0.26	0.65	4.57	< 0.01
MD Risk1 (>25th)	0.31	0.12	0.06	0.55	2.47	0.01
MD Risk2 (≤25th)	-0.18	0.15	-0.46	0.11	-1.23	0.22
Some Risk (SES)	-0.06	0.17	-0.40	0.27	-0.35	0.72
Tau^2	0.022					
I^2	30.52%					

Note. For MD Risk1, MD Risk2, and Some Risk $Q_{(3)} = 18.17$, $p < 0.01$) for the selected sample and $Q_{(2)} = 13.01$, $p < 0.01$) for the non-selected sample. CRA = concrete-representational abstract framework; MD = mathematics difficulty; SES = socio-economic status; Correspondence = counting with one-to-one correspondence; Duration = weeks of intervention.

Similar to the results of the final model including effect sizes that were dependent due more than one treatment group being compared to only one control group, the coefficients for *CRA*, *Correspondence*, and *MD Risk1* were positive for both the selected sample regression results, and the non-selected regression results. The coefficients for *Duration*, *MD Risk2*, and *Some Risk* remained negative. Across the regression results, *Duration* was significant at the $p < 0.10$ level when considered for the non-selected treatment groups, but was not significant when considered with the selected treatment groups.

Variance accounted for in Number domain studies. Table 6 summarizes the final model for the Number domain studies meta-regression results. In the subgroup of studies that included at least one Number domain skill, results indicated that *CRA*, *Correspondence* and *Numeral Identification* were significant predictors of treatment outcomes. The model was statistically significant (Q_M (3) = 31.21, $p < 0.01$; Pseudo R^2 = 67%). The coefficients for *CRA* and *Correspondence* were positive, indicating that interventions with at least one Number domain skill that included these components yielded larger treatment effects on average. In contrast, the *Numeral Identification* covariate was negative, indicating that Number domain interventions that included this skill yielded a smaller average treatment effect.

Table 6

Summary of the Number Domain Meta-Regression Model

Covariate	*B*	*SE*	95% Confidence Int. Lower	95% Confidence Int. Upper	Z-value	*p*-value
Intercept	0.56	0.14	0.29	0.82	4.11	< 0.01
CRA	0.24	0.11	0.03	0.45	2.25	0.02
Correspondence	0.41	0.10	0.21	0.61	4.02	< 0.01
Numeral Identification	-0.32	0.12	-0.55	-0.09	-2.72	< 0.01
Final Model						
Tau^2	0.037					
I^2	44.83%					

Note. *k* = 42. CRA = concrete-representational abstract framework; Correspondence = counting with one-to-one correspondence. Based on the final model, the predicted treatment effect for Number domain interventions is represented as $T_{(NUMBER\ treatment\ effect)} = 0.56 + 0.24(CRA) + 0.41(Correspondence) - 0.32(Numeral\ Identification)$

Variance accounted for in Relations domain studies. Table 7 summarizes the final model for the Relations domain studies meta-regression results. In the subgroup of studies that included at least one Relations domain skill, results indicated that *Ordinal Numbers* and *Set Comparison* were significant predictors of treatment outcomes. The model was statistically significant (Q_M (2) = 18.19, $p < 0.01$; Pseudo R^2 = 53%). The coefficients for both covariates were positive, indicating that interventions with at least one Relations domain skill that included these skills yielded larger treatment effects on average.

Table 7
Summary of the Relations Domain Meta-Regression Model

Covariate	*B*	*SE*	95% Confidence Int. Lower	Upper	Z-value	*p*-value
Intercept	0.37	0.08	0.22	0.53	4.63	< 0.01
Ordinal Numbers	0.54	0.14	0.27	0.81	3.89	< 0.01
Set Comparison	0.28	0.10	0.08	0.47	2.81	< 0.01
Final Model						
Tau^2	0.042					
I^2	48.08%					

Note. $k = 45$. Based on the final model, the predicted treatment effect for Relations domain interventions is represented as $T_{(RELATIONS\ treatment\ effect)} = 0.37 + 0.54(Ordinal\ Numbers) + 0.28(Set\ Comparison)$

Variance accounted for in Operations domain studies. Table 8 summarizes the final model for the Operations domain studies meta-regression results. In the subgroup of studies that included at least one Operations domain skill, none of the Operations skills (i.e., addition and subtraction, composing and decomposing, place value, properties of addition and subtraction, equivalence) were significant. The results indicated that *Grade* and *CRA* were significant predictors of treatment outcomes. The model was statistically significant (Q_M (3) = 20.02, $p < 0.01$; Pseudo $R^2 = 56\%$). *Preschool* served as the reference group for grade; therefore, the negative coefficients for both *Kindergarten* and *First Grade* indicate that both treatment groups from those grade levels yielded smaller effects on average compared to preschool students. The coefficient for *CRA* was positive, indicating that interventions with at least one Operations domain skill that taught skills in a *CRA* framework yielded larger treatment effects on average.

Table 8
Summary of the Operations Domain Meta-Regression Model

			95% Confidence Int.			
Covariate	*B*	*SE*	Lower	Upper	Z-value	*p*-value
Intercept	1.01	0.25	0.52	1.49	4.02	< 0.01
Kindergarten (1)	-0.37	0.25	-0.85	0.11	-1.50	0.13
First Grade (2)	-0.70	0.25	-1.20	-0.21	-2.80	< 0.01
CRA	0.21	0.13	-0.04	0.46	1.67	< 0.10
Final Model						
Tau^2	0.050					
I^2	53.31%					

Note. k = 37. Kindergarten and First Grade $Q_{(2)} = 13.30$, $p < 0.01$); CRA = concrete-representational abstract framework. Based on the final model, the predicted treatment effect for Operations domain interventions is represented as $T_{(OPERATIONS\ treatment\ effect)} = 1.01 + 0.21(CRA) - 0.37(Kindergarten) - 0.70(First\ Grade)$

Publication bias. The results for publication bias are discussed considering only the sample of studies with the two outliers removed. Figure 9 displays the funnel plot for the sample of 49 treatment effects. The plot shows fairly even distribution of effect sizes around the mean treatment effect, and does not show a clustering of effect sizes in the bottom right corner of the funnel plot, which would suggest bias toward those studies with larger effect sizes. The classic *Fail-safe N* analysis reported that an additional 4,451 studies with a null effect would need to be retrieved and included in this meta-analysis for the mean effect to be nullified. It is highly unlikely that more than 4,000 other studies exist on early numeracy interventions for preschool to first-grade students; thus, I determined that the threat of publication bias in this meta-analysis was relatively low.

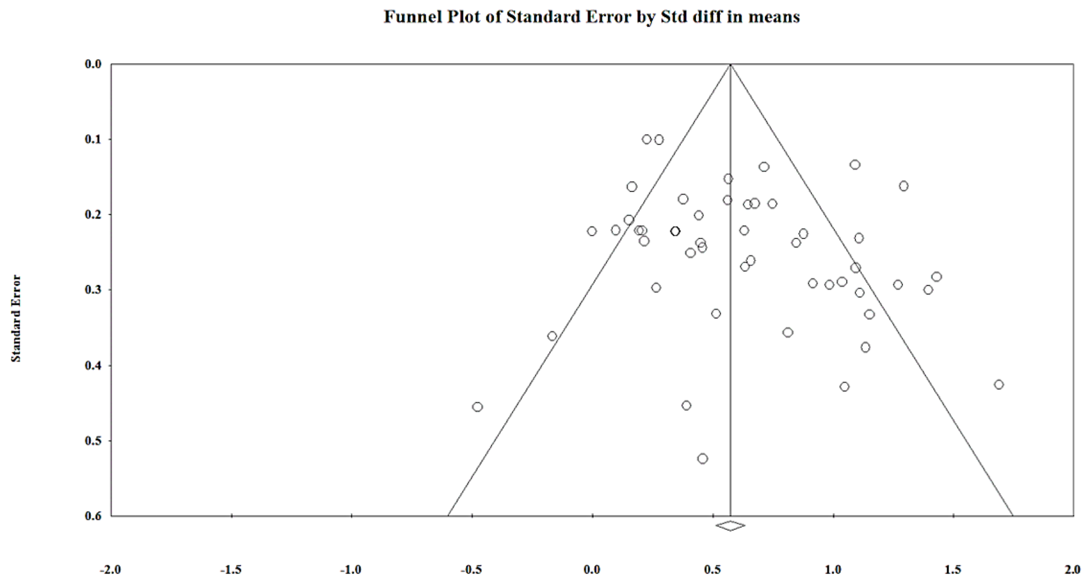

Figure 9. Funnel plot of sample with outliers removed.

Chapter V

Discussion

The purpose of this study was to provide a comprehensive review of preschool, kindergarten, and first-grade mathematics studies to estimate the effect of early numeracy interventions for all students including those with disabilities and mathematics difficulty (MD), as well as examine potential causes of variation in treatment effects. An extensive search of the literature on mathematics interventions for preschool, kindergarten, and first-grade students yielded 33 studies that met inclusion criteria for the purpose of this meta-analysis; in total, 51 treatment groups were coded to examine the effects of numeracy interventions across components such as participant characteristics, methodological features, and intervention features. In this chapter, I provide a summary and discussion of the results organized by research questions. The results are discussed only considering the sample with the two outliers removed (Clements, 1984; Mohanty & Mishra, 1994). I also compare findings from this study with results of previous syntheses. Limitations of this study are also presented, and I discuss practical implications and recommendations for future research. The chapter concludes with a summary statement of each research question and conclusion.

How Effective Are Early Numeracy Interventions?

The results of this review revealed a moderate weighted mean effect ($g = 0.63$, 95% CI [0.50, 0.73]) for numeracy interventions for preschool, kindergarten, and first-grade students. The results of the sensitivity analysis also indicated that including studies ($n = 7$) with dependent effects yielded similar results compared to including only independent effects. Moreover, the treatment groups identified as the “non-selected” groups actually yielded a lower estimated mean effect, suggesting that the inclusion of both treatment groups from studies with dependent effects may actually produce a more conservative estimate of the overall weighted mean effect. Thus, dependent treatment groups were included in this meta-analysis. The result of this meta-analysis yielded a larger average effect compared to another review that specifically examined the effects of numeracy interventions for preschool and kindergarten students (0.48; 95% CI [0.35, 0.60]; Malofeeva, 2005). The difference in results may be attributable to varying inclusion criteria and the fact that the Malofeeva (2005) literature search ceased in 2003. In contrast, the results of this meta-analysis produced a similar average treatment effect as results of other syntheses of mathematics interventions for kindergarteners to students who were 12 years old (Kroesbergen & Van Luit, 2003) and students in kindergarten through 12th grade (Gersten, Chard, et al., 2009). The results of this review are promising because 17 treatment groups (not including outliers) yielded large effects (i.e., effect sizes greater than $g = 0.80$), and the 95% confidence interval did not include zero; however, three studies did report a negative or null effect for the treatment group. Moreover, the majority of treatment groups ($k = 41$; 80.39%) surpassed the WWC criteria of 0.25 for a substantial effect (WWC, 2014). However, readers should take caution when considering 0.25 as a threshold for a substantial effect of an intervention as this review did not consider WWC study design requirements for inclusion (e.g., pretest differences greater than 0.25 for treatment and control groups).

The results suggest that early numeracy interventions successfully addressed mathematics skill deficits for participants included in these studies, including students with disabilities and MD. This finding is critical because the results of previous studies suggest that early numeracy skills measured in kindergarten and first grade are strongly related to later mathematics achievement, specifically broad mathematics achievement and computation skills (Aunola et al., 2004; Desoete, Stock, Schepens, Baeyens, & Roeyers, 2009; Friso-van den Bos et al., 2015; Locuniak & Jordan, 2008; Missall, Mercer, Martínez, & Casebeer, 2012). Because of the additive nature of mathematics (i.e., it is difficult to learn and understand fractions or algebra if a student has gaps in knowledge of other skills such as counting and simple arithmetic), early intervention is necessary when students do not possess an understanding of foundational numeracy skills. Early efforts to intervene and address gaps in early numeracy skills may prevent students from falling further behind as more advanced mathematics concepts are introduced in later grades.

The first research question also aimed to evaluate the influence of the type of outcome measure (i.e., proximal versus distal outcomes) on the overall effectiveness of the early numeracy intervention. Many treatment groups were assessed with only one mathematics outcome measure; therefore, comparing differences in treatment effects of interventions on proximal and distal outcome measures was impractical. Instead, the primary outcome measure for each study was coded as (a) proximal or distal, and (b) broad or narrow, and the treatment effect based on the type of primary outcome measure was evaluated. The primary outcome measure for most treatment groups was identified as a broad proximal measure (51%) or broad distal measure (31%). It was anticipated that many studies would include broad measures of mathematics or numeracy understanding because the majority of studies focused on interventions that emphasized more than one numeracy skill or spanned more than one domain of early numeracy. Broad measures of numeracy are aligned well to interventions that address many numeracy components, whereas results from narrow numeracy measures would not generalize well to numeracy as a broad domain of mathematics. Fewer intervention groups were administered narrow proximal measures (10%) or narrow distal measures (8%).

It was unexpected that treatment groups assessed with narrow distal measures, on average, would produce the largest treatment effect ($g = 1.16$; $SE = 0.24$), but this result represented only four treatment groups, all from the same study (Desoete & Praet, 2013). Participants were administered a subtest of a mathematics achievement measure that included items on arithmetic operations while the intervention focused on counting and comparison principles. Not including the results from Desoete and Praet (2013), results for treatment groups who were administered broad proximal measures produced the largest effect ($g = 0.76$; $SE = 0.07$). The results also indicated that norm-referenced measures yielded smaller effects ($g = 0.51$; $SE = 0.09$) than researcher developed measures ($g = 0.75$; $SE = 0.08$). Results from previous syntheses reported both similar (Gersten, Chard, et al., 2009; Xin & Zhang, 2012) and different patterns (Malofeeva, 2005) for researcher developed measures. The results of this meta-analysis were expected because researcher-developed measures are likely to be more closely aligned with the intervention content, which generally makes the measures more sensitive to changes in student achievement compared to measures that are norm-referenced. Norm-referenced measures typically include a broader scope of mathematics content (i.e., patterns, geometry; e.g., *Stanford Early School Achievement Test*) compared to the content that is included in the intervention; growth on specific numeracy interventions may not always be reflected in these measures.

Although researcher-developed measures produced a larger treatment effect on average, measures that are not norm-referenced often lack reliability and validity information which may weaken the validity of the study. Several studies that used researcher-developed measures failed to provide reliability and validity information of the measure (Kaufmann et al., 2005; Ramani & Siegler, 2008; Siegler & Ramani, 2009; Van de Rijt & Van Luit, 1998; Wilson et al., 2009). Thus, when evaluating the effect of an intervention, it may be wise to consider the degree to which the outcome measure is aligned to the intervention content, as well as the technical adequacy of measures.

Which Early Numeracy Domain was Most Investigated and Produced the Largest Effect Size?

The second research question evaluated the frequency with which interventions focused on specific early numeracy domains and the skills within each domain to determine if specific skills influenced treatment effects. The majority of interventions focused on Number and Relations skills, with slightly fewer interventions focusing on Operations skills. Most Number and Relations skills are represented in the kindergarten CCSS (NGA & CCSSO, 2010) and the preschool and kindergarten NCTM *Focal Points* (NCTM, 2006), whereas many Operations skills are not emphasized in standards and recommendations until first grade (NCTM, 2006; NGA & CCSSO, 2010). Perhaps fewer interventions addressed skills in the Operations domains because only seven studies included participants in first grade. Interventions across each of the three domains produced moderate effects. Frye et al. (2013) found moderate levels of supporting evidence for preschool and kindergarten instructional programs that focused on teaching students number and operations skills. Kroesbergen and Van Luit (2003) also evaluated the effectiveness of interventions for students in kindergarten through 12 years old that focused on preparatory mathematics (e.g., counting) and basic skills (e.g., simple addition and subtraction). The authors reported that interventions that focused on basic facts yielded larger treatment effects than interventions that focused on preparatory skills; however, both domains yielded large effects. Chodura et al. (2015) also reported a significant effect for interventions that focused on basic arithmetic competencies for students ages 6 to 12 years old.

The frequency and effect of specific skills was also assessed. Many interventions included content that addressed counting, comparison, numeral identification, and number line understanding; each of these skills also produced moderate-to-large or large effects. It is encouraging that several interventions in this meta-analysis included these skills because previous research has examined the relation between specific early numeracy skills measured in kindergarten and first grade and later mathematics achievement (Aunola et al., 2004; Desoete et al., 2009; Friso-van den Bos et al., 2015; Locuniak & Jordan, 2008; Missall et al., 2012). For example, stronger counting abilities measured in kindergarten significantly predicted higher broad mathematics achievement in second grade (Aunola et al., 2004), and first grade counting skills accounted for nearly 30% of variance in early arithmetic scores in third grade (Desoete et al., 2009). Numeral identification (Baglici, Codding, & Tryon, 2010) and quantity discrimination (Desoete, Ceulemans, De Weerdt, & Pieters, 2012) measured in kindergarten were also significant predictors of calculation abilities in first and second grades. Missall et al. (2012) also reported that quantity discrimination measured in kindergarten significantly predicted broad mathematics achievement in third grade. In order to be successful in mathematics across grade levels, students must establish foundational skills because most newly acquired mathematics skills require use and understanding of previously learned mathematics skills. For example, concepts of addition and subtraction require knowledge of counting, cardinality, and comparison. When students have gaps in mathematics knowledge, those gaps contribute to difficulty and lower performance in mathematics; therefore, early intervention that focuses on foundational numeracy concepts and whole number understanding can help prevent difficulty with learning mathematics later on.

In addition to specific early numeracy skills within the three domains, the influence of mathematics vocabulary on treatment effects was also assessed. Ten (20%) treatment groups received interventions that explicitly included a mathematics language or vocabulary component. The results of this meta-analysis suggested that studies that explicitly included mathematics vocabulary produced large effects ($g = 0.81$; $SE = 0.13$), while studies that did not include vocabulary yielded slightly smaller effects ($g = 0.58$; $SE = 0.06$). Although Frye et al. (2013) reported that there was only minimal evidence to support the incorporation of mathematics verbalizations and vocabulary during instruction for young children, the authors were optimistic in their recommendation for addressing these skills. In mathematics, students solve problems by manipulating numerals and interpreting symbols, but an understanding of mathematics requires more than knowledge of numerals and symbols. Many mathematics tests, such as state assessments, require interpretation of written language, so it is also critical for students to mastery vocabulary terms in addition to concepts and procedures.

What are the Differential Treatment Effects of Early Numeracy Interventions on Mathematics Outcomes across Categories of Coded Variables?

The third research question aimed to determine the potential causes in variation of the overall mean effect. The weighted average effect sizes for subcategories and the within- and between-groups variation was examined. Generally, the majority of variables examined reported significant heterogeneity ($p < .01$) within groups; however, not all groups reported significant heterogeneity between groups (e.g., location). This finding indicates that although there is variation within a feature of a study that was examined, variation between those features may not be present. The results of this research question are discussed separately for study information, study quality, methodological characteristics, participant characteristics, and intervention characteristics (for a discussion of dependent measures, see research question 1 above).

Study information. With regard to the location of the studies, further examination is warranted. Studies conducted outside of the U.S. produced larger effects compared to studies conducted in the U.S., suggesting that there may be something fundamentally different about these two groups of studies. For example, other countries likely use different curricula, have different expectations for students' learning, and have different teacher licensure requirements. Furthermore, many of the studies did not report qualitative data regarding ethnicity and language proficiency, so it was difficult to determine if studies outside of the U.S. had more or less of a homogeneous sample compared to studies conducted in the U.S. Although the interventions may not be substantially different, the samples in the studies may not represent the same populations.

Although I conducted an electronic search of the data from the last 35 years, the majority of studies were published in the last 10 years and more than half were conducted in the U.S. The abundance of recent literature is likely due to improved educational funding for research projects and the increased awareness of the importance of developing foundational mathematics skills at an early age (NCTM, 2006; Jordan et al., 2009; Judge & Watson, 2011). Studies that were conducted in 2012 or later had nearly identical treatment effects as those studies conducted in 2011 or earlier.

Methodological characteristics. Interestingly, none of the methodological characteristics examined in this meta-analysis produced significant between-groups *Q* test results. I predicted that interventions that supplemented the regular mathematics instruction would yield significantly larger effects due to treatment groups receiving more mathematics instruction, compared to interventions that supplanted instruction (indicating a time control for the treatment and control groups). However, a closer look at each group of studies (supplanted full mathematics instruction, supplanted part of mathematics instruction, supplemented, and not reported) revealed that only "not reported" and "supplanted full" interventions included preschool participants, and when considering grade alone, preschool treatment groups yielded larger effects than both kindergarten and first-grade treatment groups. Moreover, the category "not reported" included six of the seven preschool studies; this raises some questions regarding the nature of preschool general mathematics instruction in this sample of studies. When authors do not report, for example, the amount of regular mathematics instruction students receive at this grade level, the quality of the control condition, and if the intervention supplanted or supplemented the regular instruction, it is difficult to identify potential influences of larger treatment effects for preschool students who are receiving numeracy interventions.

I utilized codes for study quality and confidence to increase assurance in the information gathered from studies and the codes that were applied to each treatment group. To gain a sense of the quality of the study, I specified if a study provided information on participant and intervention agent demographics, fidelity of implementation, features of the control condition, and attrition. There was very little variation in treatment effects when considering how many of the five indicators studies contained. Regarding confidence, studies that received a rating of “very confident” produced larger treatment effects than studies that received a rating of “mostly confident” regarding the information presented (this rating differs from study quality in that the accuracy of information presented was evaluated); however, this difference was not statistically significant. Although this finding is positive considering that even studies that do not provide information associated with study quality (Gersten et al., 2005) and studies that include some discrepancies in the information provided still report effective intervention outcomes, it does raise questions regarding the generalizability of findings and confidence in replicating intervention effects in classroom settings. For example, it is difficult to know if findings will generalize to other contexts if information regarding the participants or intervention agent is not provided. Practitioners and researchers would benefit from more studies that provide detailed and accurate information to allow for generalizability and replication.

It is also worth restating that there was no statistically significant between-groups variation between treatment groups identified as independent and dependent. Dependent studies on average actually yielded a smaller treatment effect, which suggests including all of the dependent treatment groups in the total sample resulted in a more conservative estimate of the overall weighted random mean effect.

Participant characteristics. Most studies did not provide enough information (or did not provide disaggregated results) on gender, race, free or reduced lunch (FRL), or English learner (EL) status of participants to examine outcomes as a function of those characteristics. Thus, the results of this meta-analysis make it difficult to establish important patterns related to participants' characteristics and the effectiveness of early numeracy interventions. English learners and students from low-income families traditionally perform below peers on measures of academic achievement (National Center for Educational Statistics, 2013a), so it is critical to understand the differential effects of numeracy interventions for these students in order to address the mathematics achievement gap that exists for many students as early as kindergarten.

Conversely, studies did provide sufficient information regarding participants' grade or age. More than half of studies focused on participants in kindergarten, with less representation of preschool and first-grade participants; it was not surprising that most numeracy interventions focused on kindergarten students considering the kindergarten CCSS in mathematics emphasizes early numeracy skills (i.e., counting and cardinality are not addressed in first-grade CCSS), while first-grade standards begin to focus on the importance of fluency with basic operations. Interventions with preschool and kindergarten participants produced large and moderate-to-large effects, respectively, while interventions with first-grade participants yielded small effects; difference in treatment effects by grade were statistically significant. Perhaps the larger treatment effects for preschool participants is related to the fact that most preschool studies did not report the nature of the treatment condition, and most of the preschool studies described the nature of the control group as circle or center time, and free play. With little detail about the nature of the treatment and control, it was difficult to determine if preschool participants in these studies were actually receiving regular mathematics instruction. The large effects for preschool treatment groups could be explained by the lack of exposure to regular mathematics instruction by the control group. The NRC Committee on Early Childhood Mathematics examined the research on early mathematics and uncovered a shortage of opportunities to learn mathematics in school, especially when compared to the number of opportunities to learn reading. For example, observations conducted by La Paro and colleagues (2009) revealed that language and literacy instruction consumed 28% of time in kindergarten classrooms, while other content areas (e.g., mathematics) consumed less than 10% of instructional time. If students in preschool have fewer opportunities to learn mathematics they likely have more room to grow.

More than half of treatment groups were identified as at-risk for MD using a mathematics screening measure. The most common method for identifying participants as having MD was performance below a specified criterion on a mathematics measure, and as expected there was variability in author-specified criterion of MD even when using a cut score (e.g., performance below the 25th percentile, 40th percentile, between the 25th and 50th percentile, 1.5 *SDs* below the mean), because, generally, there is a lack of consensus regarding the criteria to identifying participants with MD (Nelson & Powell, in press). Only three treatment groups were identified as being at-risk for MD due to having a documented disability or being referred for special education services; this was anticipated due to the unlikelihood that students are identified for special education services before first grade (Holt, McGrath, & Herring, 2007). Though fewer studies examined the effects of preschool interventions, the most common method to identify potential academic risk for preschool students was participation in Head Start; while the one remaining preschool study with participants identified as having some academic risk that did not use this method was conducted outside of the U.S. where Head Start participation is not applicable.

Variation between participants' risk status groups was not significant; although it is worth noting that average treatment effects did vary slightly according to risk status. Interestingly, effect sizes were slightly higher on average for students who were identified as at-risk for MD according to (a) low socioeconomic status (SES), and (b) performance below an author-specified criterion on a mathematics screener, but performance greater than the 25th percentile (e.g., 40th percentile). In contrast, students identified as at-risk for MD according to performance below the 25th percentile yielded smaller effects on average, though still moderate. Treatment groups with students who were not screened for MD and were considered typically achieving yielded moderate effects. So, when only considering participant risk status without considering other variables, students who were at a moderate level of risk appeared to perform slightly better than students who were not screened for MD and those who were identified as at-risk for MD with greater deficits. It is also interesting is that of the three treatment groups that had negligible or negative treatment effects, two were typically achieving or high achieving students. These results may suggest a ceiling effect on outcome measures if these students were already performing average or above average compared to other peers.

The results for students with MD (with performance above the 25th percentile) are similar to findings of other reviews that reported mathematics interventions to be effective for elementary students with special needs ($d = 0.62$; Kroesbergen & Van Luit, 2003) and students with learning disabilities ($g = 0.63$; Gersten, Chard, et al., 2009). However, there was a wide range of variability in treatment effects for groups with MD (g range = 0.19 to 1.64) and students with disabilities (g range = 0.43 to 1.02). The variable results for students identified in different MD categories may be attributed to the fact that initial student performance and growth in mathematics varies as a function of the cutoff criterion that is used to define students with MD. For example, Murphy, Mazzocco, Hanich, and Early (2007) reported that students who performed below the 10th percentile on a measure of mathematics achievement had significantly slower rates of growth than students who performed between the 11th and 25th percentile and students who performed above the 25th percentile. Furthermore, the authors did not observe differences in growth between students in the 11th – 25th percentile group and the greater than 25th percentile group. These results suggest there may be distinct differences between students who perform below the 10th percentile and those who perform between the 11th and 25th percentile. The effectiveness of a numeracy intervention for students with MD may be influenced by the degree to which these students exhibit deficits.

Regardless of the lack of statistically significant findings on the effectiveness of early numeracy interventions for the lowest performing students (students who score below the 25th percentile) included in this review, the positive moderate findings are encouraging, especially in light of these students' history of failure in mathematics and the achievement gap compared to typically performing peers. As previously discussed, children who enter school with deficient mathematics skills continue to fall further and further behind typically achieving kindergarten peers throughout elementary school (e.g., Morgan et al., 2009). Several of the studies included in this synthesis reported large effects (i.e., effect sizes greater than 0.80) for students with MD or students with disabilities (e.g., Clarke et al., 2014; Desoete & Praet, 2013; Schopman & Van Luit, 1996; Sood & Jitendra, 2011), which suggests that preschool through first grade is an advantageous and critical time to implement mathematics interventions to adequately address the gap between low and high entry mathematics performance for young students.

Intervention characteristics. The results indicated that only instructional arrangement yielded significant between-groups variation for the intervention characteristics examined in this meta-analysis; interventions with treatment groups who received instruction in small groups (3–5 students), in flexible grouping, and one-to-one, yielded moderate-to-large and large effects (average g range = 0.57–1.28). In contrast, treatment groups who received instruction using peer-assisted learning strategies produced small effects on average ($g = 0.25$; $SE = 0.13$), though these results were within the general range of what some previous reviews reported. Gersten et al. (2009) found a small effect ($g = 0.14$) for peer-assisted instruction, whereas Kunsch, Jitendra, and Sood (2007) found a moderate effect ($d = 0.47$) for peer-assisted mathematics strategies for students with learning disabilities and difficulties. However, Baker et al. (2002) and

Malofeeva (2005) found relatively larger effects ($d = 0.62$; $g = 0.71$, respectively) for peer-assisted instruction in mathematics. The differences between the results of this study and previous studies may be attributable to varying inclusion criteria for participants (e.g., age, risk-status).

Intervention agent and length of intervention were also examined. The between-groups variation that was reported for intervention agent was not significant, treatment groups yielded similarly moderate-to-large effects across intervention agent type (average g range = 0.56–0.76). Within this sample of early numeracy studies, the effect of the intervention did not appear to be substantially related to the intervention agent; this result may be due to the large number of treatment groups who received instruction with scripted lessons ($k = 36$). Scripted lessons may result in higher rates of fidelity of implementation and control for any effects of intervention agent. With regard to length of treatment, similar moderate-to-large effects (average g range = 0.61–0.64) were found for interventions that lasted fewer than 8 weeks and interventions that lasted longer than 8 weeks. Kroesbergen and Van Luit (2003) reported that duration of the intervention in their review was a significant predictor of the effect of the intervention and the longer an intervention lasted, the less effect it had. Meanwhile, Xin and Jitendra (1999) reported the largest treatment effects for interventions lasting more than 1 month. Although the results of this meta-analysis do not clarify the degree to which the length of an intervention influences the effectiveness of the treatment for early numeracy interventions, reviews that have examined length of treatment have arbitrarily selected cut-offs for weeks or months of intervention without consistency across studies. Therefore, it is difficult to compare and generalize results of this meta-analysis.

Instructional features. Several instructional features were considered in this meta-analysis, such as components of explicit and systematic instruction, scripted lessons, concrete-representational-abstract (CRA) framework, and game-based instruction. Though not to a statistically significant degree, interventions that included a CRA framework for teaching mathematics concepts yielded larger effects. This finding was not surprising as many previous reviews have reported similar findings that using multiple mathematics representations (such as in a CRA framework) results in greater outcomes for students, and that using multiple representations has been identified as an evidence-based practice in mathematics (Bouck, Satsangi, & Park, in press; Gersten, Beckmann, et al, 2009; Gersten, Chard, et al., 2009; Jitendra, Nelson, Pulles, Kiss, & Houseworth, 2016; Mononen et al., 2015). What was surprising, however, was interventions that included "concrete manipulatives" or "pictorial representations" in the absence of a CRA framework did not yield larger treatment effects. This result underscores the idea that representations alone may not result in greater intervention outcomes. Many factors may influence the effectiveness of multiple representations, including using concrete and pictorial representations in a graduated sequence, teachers modeling of appropriate use of representations, or opportunities for students to practice with representations (as opposed to only allowing teachers to model with representations). Further investigation of factors that may influence the effectiveness of concrete manipulatives and pictorial representations is needed. Authors of future reviews of numeracy interventions may consider coding studies for more detailed information related to multiple representations, such as the specific type of representation and frequency of use.

Interventions that included at least one component of explicit and systematic instruction produced smaller treatment effects on average compared to studies that did not report including at least one component of explicit and systematic instruction. This result is different than what previous reviews have reported regarding explicit and systematic instruction (Baker et al., 2002; Gersten, Beckman, et al., 2009; Gersten, Chard, et al., 2009; Mononen et al., 2015). This result should be considered with caution and does not suggest that explicit and systematic instruction is not effective; only two treatment groups did not report including at least one component of explicit and systematic instruction and even though the average treatment effect for studies that did include components was smaller, the average treatment effect for explicit and systematic instruction was still moderate-to-large ($g = 0.62$; $SE = 0.06$). Furthermore, the coding procedures I used in this meta-analysis could have played a role in this result, as I grouped many components of explicit and systematic instruction together in one code. When considering specific components of explicit instruction individually, such as teacher modeling, guided and independent practice opportunities, and providing corrective feedback, the results indicate very similar moderate effects for interventions that included those explicit and systematic instructional features versus interventions that did not include those features. The lack of detail provided by authors regarding instructional features may not allow for the most accurate coding for each intervention component; in other words, even though authors may not have described instructional features or techniques used by intervention agents, it is not safe to assume that intervention agents did not use modeling, guided practice, corrective feedback, and so on. Future research on interventions would benefit from more detailed descriptions of instructional features to allow for better replication of intervention programs and treatment effects.

Finally, interventions that were administered in a game-based format or included games generally produced moderate-to-large effects (average *g* range = 0.49–0.72), though only "included games" produced statistically significant between groups variation in favor of those treatment groups that received interventions that did not include games. Regardless of whether the intervention group received instruction with games or in a game-based format, interventions were effective. This result may be worth evaluating further in future research to determine which types of games yield larger treatment effects or if the instructional format influences the outcomes of interventions with games (e.g., is playing a game with an interventionist more effective than playing a game with a peer).

Which Variables Account for the Most Between-Studies Variance?

The aim of the final research question was to determine which variables explain a large portion of the variance between-studies that is not explained by sampling error. First, I examined the total sample of studies, followed by subgroups of studies that addressed content in each of the early numeracy domains, Number, Relations, and Operations. In selecting each final model, I considered previous research, the practical significance of the variable, and the results of the first three research questions.

Total sample. The final model selected for the total sample of studies included the predictors of CRA, intervention duration, risk status of participants, and the inclusion of counting with one-to-one correspondence in the intervention content. Similar to the results of the other research questions and previous research (Gersten, Beckmann, et al, 2009; Mononen et al., 2015), interventions that presented content in a CRA framework resulted in post-test gains in mathematics. Although Malofeeva (2005) also evaluated the effectiveness of early numeracy interventions, the author reported that only instruction on measurement concepts was a significant predictor of treatment outcomes. In contrast, the results of the final model of this meta-analysis indicate that counting with one-to-one correspondence had a significant and positive effect on the magnitude of the treatment outcome. Students in early elementary grades with MD often demonstrate lower performance on counting tasks and longitudinal research reports that students with low counting skills in kindergarten tend to exhibit difficulty with other mathematics skills later on such as arithmetic (Desoete & Grégoire, 2006; Stock, Desoete, & Roeyers, 2010). The results of this meta-analysis add to the research base on the importance of developing counting skills and the significance of including counting skills in early numeracy interventions. However, it is also worth noting that although no other numeracy skills (e.g., numeral identification, cardinality, composing) were included in the final meta-regression model of the full sample, other numeracy skills across all domains are still necessary for students' development of solid whole number understanding. It is also wise to consider that not all students who need an early numeracy intervention have deficits in the area of one-on-one correspondence or similar skills. Practitioners may consider screening for specific deficits in early numeracy skills before determining which skills to include in an intervention.

When controlling for the other variables in the model, duration also had a significant and negative effect on the treatment outcome, indicating that longer interventions (greater than 8 weeks) actually yielded significantly smaller effects. This is similar to the results reported by Kroesbergen and Van Luit (2003). When examining only the duration of the intervention, the results of research question 3 suggested that both (a) interventions lasting 8 weeks or less and (b) interventions lasting longer than 8 weeks produced similar moderate-to-large effects; however, the results of the meta-regression indicate that when controlling for other intervention features such as content and risk status, interventions that are longer in duration may actually be less effective. It may be important for educators to implement progress monitoring to track student progress toward intervention goals to determine if more weeks of intervention are needed. Furthermore, duration in this meta-regression was considered only as the number of weeks of intervention a student received. Other factors related to time, such as the number of intervention sessions per week and the length of each intervention sessions, may be considered in future reviews of early numeracy interventions.

Finally, when examining the participant risk-status in previous research questions, results indicated moderate effects for participants with low academic risk and risk for MD with performance below the 25th percentile on a screener, and large effects for participants with some risk (SES) and risk for MD with performance above the 25th percentile on a screener; between-groups variation was not statistically significant. However, the results of the meta-regression indicate that when controlling for other variables, participant risk-status (overall) was a significant predictor of treatment outcome and a slightly different pattern emerged. When controlling for other variables, interventions that included participants with some risk (low SES) actually yielded effects that were less effective compared to the reference group (typically-achieving students). Future examination of risk for academic failure in mathematics may be considered, as only a few studies identified students as at-risk with this criteria, and criteria was not consistent across these studies. For example, preschool studies identified risk with participation in Head Start (which is based on income requirements), other studies identified SES risk with average FRL percentages (meaning that there is a chance some students in that sample were not from low SES), and one study was conducted outside the U.S. where poverty rates and criteria may not align to U.S. rates and criteria.

Interventions were also less effective for students with MD identified with more restrictive criteria (i.e., performance at or below the 25th percentile) compared to typically achieving students. The results of one of the studies included in this meta-analysis also reported that even when the average effectiveness of an intervention is reported as moderate, the subgroup performances of students with more severe MD may in fact be much lower than both the average effect and the effect with less severe MD (Toll & Van Luit, 2012). Moreover, the results of this meta-regression also suggest that students with at least some risk of MD (i.e., performance above the 25th percentile), but not severe risk, performed better than typically-developing peers. This result may be attributed to the fact that students with at least some risk had more opportunity to improve mathematics performance during the intervention period, while there may have been ceiling effects for those students identified as typically achieving. In summary, it is imperative that researchers and practitioners acknowledge that students of varying mathematics abilities may represent very different groups of students, each with unique instructional needs that may require different forms of intervention (e.g., group size).

Number domain. In the subgroup of studies that included at least one Number domain skill, results indicated that similar to the total sample of students, both CRA and counting with one-to-one correspondence were significant predictors of treatment outcomes. When considering Number domain interventions only, it was surprising that more Number domain skills did not significantly and positively add to the final meta-regression model; however, approximately 70% of interventions that included counting with one-to-one correspondence also included counting sequence and cardinality skills. With the overlap in counting skills, future research may consider evaluating these skills further through examination of their effectiveness as measured by narrow or single-skill measures.

Numeral identification was also a negative significant predictor of treatment outcomes. Although the results of this meta-regression suggest that including numeral identification skill in early numeracy interventions is actually harmful for treatment effects, after further investigation of the interventions that included this skill, a pattern related to the location of the study emerged. Of the 29 treatment groups that received an intervention that included numeral identification skills, 83% percent were studies conducted in the U.S. In contrast, of the studies that did not include numeral identification, only 36% were conducted in the U.S. Studies conducted in the U.S. produced significantly smaller effects compared to studies conducted outside the U.S.; the negative coefficient for numeral identification in the meta-regression may be reflective of an interaction with the location of the study, as opposed to the inclusion of the skill in the intervention. This result may be due to the fact that outside of the U.S. different core mathematics curricula are used, in addition to different intervention packages; therefore, future research may investigate effects of interventions based on the study location or future reviews may consider reporting results separately for different locations in order to increase the generalizability of the results.

Relations domain. In the subgroup of studies that included at least one Relations domain skill, results indicated that ordinal numbers and set comparison were significant and positive predictors of treatment outcomes. Regarding the significance of ordinal numbers, only seven intervention groups received instruction in this skill (all treatment groups were preschool and kindergarten). Though a small number of intervention groups makes it difficult to provide recommendations regarding the inclusion of this ordinal number skill in early numeracy interventions, practitioners may want to consider the distinctiveness of the skill compared to other numeracy skills. Ordinal number skill requires quantification (e.g., the quantity 1 is related to the ordinal position “first”) and comparison (e.g., the second person in a grocery line is closer to the checkout than the fourth person in line) in ways that may be different than how traditional counting and comparison tasks are presented.

Operations domain. Unlike the meta-regression results for the subgroup of studies from the Numbers and Relations domains, the meta-regression for the Operations domain reported that none of the Operations skills were significant sources of variation in treatment effects. Meanwhile, other reviews have reported on the significant treatment effect for interventions that focus on operations skills (Chodura et al., 2015; Kroesbergen & Van Luit, 2003). The results of this review should be considered with caution as the results may be due to the fact that a smaller number of studies addressed Operations domain skills in interventions. Very few interventions specifically targeted first-grade students, and most of the Operations early numeracy skills represent first-grade content (NGA & CCSSO, 2010; NCTM, 2006). When interventions did include Operations domain skills, they also addressed fewer skills on average compared to Number and Relations skills.

Developmental progressions of mathematical development indicate the importance of Operations skills such as composing and decomposing and place value. Composing and decomposing skills are pre-requisites for mastering basic fact fluency, and mastering fact fluency allows students to expend more cognitive energy on more complex mathematics skills such as word-problem solving. Furthermore, composing and decomposing skills can be transferred to complex addition and subtraction problems that may not be solved as easily with direct retrieval (Geary, Hoard, Nugent, & Bailey, 2012). Understanding of place value becomes increasingly important as students encounter more complex computation such as multi-digit multiplication and division. Furthermore, a firm understanding of whole number operations is needed before students can master similar skills with rational numbers (Hansen et al., 2015). Finally, longitudinal research indicates that students with MD consistently display difficulties with computation, displaying correct fact retrieval, and place value concepts (Andersson, 2010; Geary, Hamson, & Hoard, 2000; Jordan & Hanich, 2003). Thus, the importance of intervening on skills related to Operations domain skills should not be overlooked due to the lack of significant results in this meta-analysis.

Similar to other results in this meta-analysis, CRA was also a significant predictor of treatment effects for studies that addressed Operations skills; grade was also a significant predictor. The negative coefficients for both kindergarten and first grade treatment groups in the meta-regression indicate that intervention groups with students from those grade levels yielded smaller effects on average compared to preschool students; however, in examining the data further, there was a pattern between the number of Operations skills addressed and grade level. Preschool treatment groups in this meta-analysis resulted in larger treatment effects on average, and they also only included an average of 0.50 Operations skills; kindergarten treatment groups yielded the next largest average treatment effect, and included an average of 2.10 Operations domain skills; first grade intervention groups reported the smallest treatment effect in this study, and included an average of 2.70 Operations skills. Operations skills in preschool may look very different in practice than they do in first-grade; therefore, future research reviews may consider coding for different levels (difficulty) of Operations skills, how Operations skills were taught, what materials were used for each type of skill, and so on, in order to examine the sources of variation in treatment effects for Operations interventions across multiple grade levels.

Limitations

As with any review, there are limitations with conducting a meta-analysis. First, meta-analyses are often limited by the "file drawer problem." Although the majority of the studies in this review had an effect size greater than zero, inferences may be somewhat biased if studies with non-significant or negative effects exist but were not included in this meta-analysis (Borenstein et al., 2009). However, the results of the publication bias analysis indicate that the chance that the mean treatment effect is actually zero is small considering the large number of studies that would need to be located with null effects. Future researchers conducting reviews or meta-analyses may help reduce publication bias by conducting exhaustive searches of the literature that include non-peer-reviewed sources, as I did in the meta-analysis. Second, a common criticism of meta-analyses is that the summary effect represents "apples and oranges." Given the wide variety of participants included in this meta-analysis (i.e., participants were identified as at risk for MD with varying criteria) and the differences in how intervention effectiveness was measured, it is worth considering the limitation that the summary effect reported for this meta-analysis may ignore important differences across studies. To address this limitation of meta-analyses, I coded each study comprehensively on features that may have differed and were potential sources of variation in treatment effects. Although the studies included in this meta-analysis did differ on important features, all of the interventions focused on improving the early numeracy skills for preschool to first-grade participants and I coded the articles to determine *how* the differences between studies influenced the summary effect. Future researchers who conduct meta-analyses may also address the oranges and apples argument by carefully considering inclusion criteria for the meta-analysis, identify potential outliers, and code studies extensively on features that differ across studies, as I did in this meta-analysis.

Another limitation specific to this synthesis was that the small number of studies that met inclusion for preschool and first-grade participants limits the inferences that can be made regarding features of interventions found effective in improving early numeracy skills for students at these grade levels. This limitation may be addressed in future reviews with an updated literature search as well as expanding search and inclusion criteria to include mathematics interventions that were excluded from this review (e.g., studies that did not disaggregate outcome data by grade level).

Many of the studies included did not report enough information or results for different populations of students (e.g., race, special education, ELs). The lack of information about participant characteristics reduces the degree to which generalizations can be made about the effectiveness of early numeracy interventions for populations that are generally more at-risk for academic difficulties, including students with MD or those who are from families with low SES. Many studies also lacked information specific to the intervention, such as providing a quantitative measurement of treatment fidelity, indicating whether the intervention was a replacement or an addition to the regular mathematics curriculum, and describing instructional features and specific intervention content in detail. Research indicates, overall, that interventions implemented with greater fidelity have better outcomes but that the optimal threshold of implementation integrity may differ depending on different features within the same intervention or for different interventions altogether (Hagermoser Sanetti & Kratochwill, 2009); procedures for measuring implementation fidelity and degree of fidelity were not available for all studies. Finally, the variability in the detail provided in studies regarding early numeracy content and instructional features may limit the scope for determining what features specifically make early numeracy interventions effective.

Future Research Directions

The results of this meta-analysis suggest that preschool, kindergarten, and first-grade numeracy interventions are effective in improving students' mathematics performance, including students with disabilities and those at-risk for MD. Further investigation, however, is needed to determine what additional sample characteristics are associated with effective interventions, and if interactions between any of the variables explain variation in the treatment effects. Future researchers should provide adequate sample characteristic information, as well as disaggregated results (i.e., means and *SD*s) for at-risk populations in primary studies, in order to generalize treatment outcomes. Although many studies included participants who were at-risk for MD, studies with other student populations did not always discuss whether or not participants in that study *may* have been at-risk (i.e., in this meta-analysis, typically achieving referred to any group of students not screened for difficulty, but lack of screening does not guarantee that no students were in fact, at risk). Future intervention studies may also consider screening all students for risk of MD and providing disaggregated results for specific groups of students (i.e., students with MD, ELs).

Future research would also benefit from more detailed descriptions and examples of lesson plans (i.e., domains addressed, specific mathematics skills, type of instructional format) for both the treatment and control conditions. It is difficult to determine the effectiveness of an intervention without also considering the instructional features and content addressed in the control condition; furthermore, it is difficult to understand the degree to which students' performance in mathematics improved if information regarding whether the intervention added to or replaced the general curriculum was not provided. In addition, future research could also provide more information regarding the level of implementation fidelity for different features of the intervention and clear descriptions of how fidelity data were gathered.

Future syntheses may also amend inclusion criteria to include a larger number of studies in order to examine additional variables and relationships between variables not examined in this synthesis. These may include training requirements for intervention agents and an analysis of mathematical representations and intervention materials.

Implications for Practice

Despite the positive results of this meta-analysis, the fact remains that many students exhibit chronic low achievement in mathematics throughout school (Duncan et al., 2007; Judge & Watson, 2011). Researchers and practitioners may take the results of this review to guide the design and implementation of future numeracy interventions that may address difficulties with learning mathematics as early as preschool. For example, schools and practitioners may benefit from results that suggest that brief interventions are successful in improving students' mathematics achievement, keeping in mind that interventions do not need to be long in duration to be effective. Although results suggest that favorable instructional features of these interventions include brief duration, some students may not respond sufficiently to brief interventions and may require longer, more intensive interventions that supplement regular mathematics instruction. Fortunately, longer interventions were also found to be effective in improving students' mathematics achievement. Also, staff other than teachers may also adequately implement effective mathematics interventions. Schools may decide to safeguard teacher resources and time by using paraprofessionals, pre-service teachers, and volunteers as intervention agents for struggling learners. At the same time, it is also important to consider the need for professional development around intervention programs and early numeracy for staff who are tasked with administering interventions. Although the results of this meta-analysis reported that interventions were effective regardless of the type of intervention agent, this meta-analysis did not specifically examine the degree to which intervention agents were trained or the type of training they received. Staff who do not have experience with administering intervention programs or do not have as much experience working with students who are at-risk for academic failure may require more professional development opportunities and oversight than certified and experienced teachers.

The results of this meta-analysis also support the research finding of the importance of teaching mathematics concepts to students using a CRA framework; furthermore, this finding can be generalized beyond teaching numeracy concepts to include other areas of mathematics, such as computation. Even when intervention lessons are scripted and do not call for the use of concrete manipulatives or visual representations, students may still benefit from the inclusion of multiple representations in order to gain conceptual understanding of numeracy skills. Teachers may look for opportunities to extend intervention content to include multiple representations or re-teaching concepts using the CRA framework. This finding also has implications for training and resources as schools may need to purchase materials and provide teachers with training to support the use of multiple representations in early mathematics instruction.

Although the results of the meta-regression did reveal a significant effect of many of the specific numeracy skills across the Number, Relations, and Operations domain, practitioners should not dismiss the importance of each of these skills. In other words, even though counting with correspondence consistently resulted in significant treatment effects compared to other numeracy skills, practitioners should not *only* focus on this skill. These results may expose the fact that more research needs to be conducted on interventions that address other numeracy skills and the fact that many intervention studies did not provide enough detailed information regarding the skills that are addressed by intervention lessons.

Moreover, teachers and interventionists may want to consider the time of the year and students' prior knowledge before they determine the most appropriate numeracy skills to include in the interventions, because young children typically acquire early numeracy skills in a hierarchical manner (i.e., Number skills are typically learned before Operations skills; Purpura et al., 2013). Most studies failed to report the time of year interventions took place and this may have an effect on how many domains are addressed and the sequence in which skills are addressed. This may have implications for practice and future research. Practitioners and researchers may want to be careful in selecting or designing numeracy interventions for students based on the skills and experiences students have already been exposed to both inside and outside of the classroom. For example, if students have already mastered the principles of counting and related skills in the *Number* domain, it will be less important to select or design an intervention with features that address counting compared to skills in other domains. This could be especially important for schools and teachers who are challenged with time and resource constraints to implement interventions.

Finally, although the main purpose of this meta-analysis did not focus on the importance of mathematics vocabulary instruction, the results of this study provide some evidence of the importance of incorporating this skill in early numeracy instruction with young students. Teachers may naturally incorporate mathematics vocabulary instruction during both core and intervention time without needing to purchase additional materials or resources. Research supports using explicit instruction to teach mathematics vocabulary (Bay-Williams & Livers, 2009; Monroe & Orme, 2002), including with students who experience difficulty with learning mathematics (Riccomini, Smith, Hughes, & Fries, 2015). Other options that teachers may use to teach mathematics vocabulary instruction include mnemonic strategies and game-based instruction (Riccomini et al., 2015).

Summary and Conclusion

This meta-analysis aimed to determine the effectiveness of and effective features of early numeracy interventions for preschool, kindergarten, and first-grade students. After an extensive literature search, 33 studies were identified for inclusion with a total of 51 treatment groups. Two of these studies were eventually identified as outliers. This meta-analysis examined four research questions.

The goal of research question 1 was to determine the overall effectiveness of early numeracy interventions for preschool, kindergarten, and first grade students. The results of this meta-analysis reported a moderate weighted mean effect (g = 0.63, 95% CI [0.50, 0.73]), and 17 treatment groups yielded large effects for numeracy interventions for preschool, kindergarten, and first-grade students.

The goal of research question 2 was to identify which early numeracy domain was most investigated and which domain produced the largest effect size. Most interventions focused on Number and Relations skills, with fewer interventions focusing on Operations skills; in addition, interventions in each domain yielded similarly moderate treatment effects.

Research question 3 examined variation within and across subcategories of coded variables, such as participant characteristics and intervention features. Interventions with preschool and kindergarten participants produced large and moderate-to-large effects, respectively, while interventions with first-grade participants yielded small effects. More than half of all treatment groups were identified as at-risk for mathematics difficulty, either by performance on a screening measure or risk determined by low socioeconomic status. On average, interventions that presented content in a concrete-representational-abstract (CRA) framework yielded larger effects.

The aim of research question 4 was to determine which variables were sources of between-studies variance for the total sample, and for each early numeracy domain separately. The final model for the total sample of studies reported larger predicted treatment outcomes for interventions that were taught in a CRA framework, included content that focused on one-to-one correspondence, were 8 weeks or less in duration, and included participants who were at some risk of mathematics difficulty as determined by performance on a mathematics screener above the 25th percentile. The final model for the subgroup of interventions in the Number domain also predicted larger treatment effects for interventions that were taught in a CRA framework and included content that focused on one-to-one correspondence, while the inclusion of numeral identification skills predicted smaller treatment effects. The final model for the subgroup of interventions in the Relations domain predicted larger treatment effects for interventions that included content focused on ordinal numbers and set comparison. The final model for the subgroup of interventions in the Operations domain did not include any Operations early numeracy skills; however, the results predicted larger treatment effects for interventions that were taught in a CRA framework and smaller effects for kindergarten and first-grade participants.

The results of this meta-analysis add to the evidence base regarding the effectiveness of mathematics interventions by specifically focusing on early numeracy content for young students, including students with disabilities and those at-risk for MD. Previous reviews have focused on other areas of mathematics including word problem solving and computation. Reviews of broad mathematics have not explicitly focused on participants from preschool to first grade, when early numeracy skills are typically taught to students. Generally, students who received early numeracy interventions experienced greater gains in mathematics performance compared to peers who did not receive a similar type of instruction. Considering that performance as early as kindergarten-entry predicts later mathematics performance, and students who enter school with poor mathematical understanding tend to remain behind their peers, the early years of school reflect a prime window to address the gap that exists between low and high performing students in foundational early numeracy skills.

References

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

Alloway, T. P. (2007). *Automated working memory assessment* (AWMA). London, UK: Pearson Assessment.

Andersson, U. (2010). Skill development in different components of arithmetic and basic cognitive functions: Findings from a 3-year longitudinal study of children with different types of learning difficulties. *Journal of Educational Psychology, 102*, 115–134. doi:10.1037/a0016838

Aunola, K., Leskinen, E., Lerkkanen, M. K., & Nurmi, J. E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology, 96*, 699–713. doi: 10.1037/0022-0663.96.4.699

Baglici, S. P., Codding, R., & Tryon, G. (2010). Extending the research on the tests of early numeracy: Longitudinal analyses over two school years. *Assessment for Effective Intervention, 35*, 89–102. doi:10.1177/1534508409346053

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

Barnett, W. S., Carolan, M. E., Squires, J. H., & Clarke-Brown, K. (2014). *State of Preschool 2013: First Look* (NCES 2014-078). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch.

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2012). Fostering at-risk kindergarten children's number sense. *Cognition and Instruction*, *30*, 435–470. doi: 10.1080/07370008.2012.720152

Baroody, A. J., Eiland, M. D., Purpura, D. J., & Reid, E. E. (2013). Can computer-assisted discovery learning foster first graders' fluency with the most basic addition combinations? *American Educational Research Journal, 50*, 533–573. doi: 10.3102/0002831212473349

Bay-Williams, J. M, & Livers, S. (2009). Supporting math vocabulary acquisition. *Teaching Children Mathematics, 16*, 238−245.

Benson, G. (2013). The effects of cover, copy, and compare, performance feedback and rewards on the mathematical calculation skills of students identified with math difficulty, (*Unpublished doctoral dissertation*). The City University of New York, New York City, New York.

Berch, D. B., & Mazzocco, M. M. (Eds.) (2008). *Why is math so hard for some children?: The nature and origins of mathematical learning difficulties and disabilities*. Baltimore, MD: Paul H. Brookes.

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

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Hoboken, NJ: John Wiley & Sons, Ltd.

Bouck, E. C., Satsangi, R., & Park, J. (In press). The concrete-representational-abstract approach for students with learning disabilities: An evidence-based practice synthesis. *Remedial and Special Education.* Advance online publication. doi: 10.1177/0741932517721712

Bryant, D. P., Bryant, B. R., Gersten, R. M., Scammacca, N. N., & Chavez, M. M. (2008). Mathematics intervention for first- and second-grade students with mathematics difficulties: The effects of tier 2 intervention delivered as booster lessons. *Remedial and Special Education, 29*, 20–32. doi: 10.1177/0741932507309712

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

*Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children, 78,* 7–23. doi: 10.1177/001440291107800101

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

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

Center for the Study of Mathematics Curriculum. (2004). *Summary of An agenda for* *action: Recommendations for school mathematics of the 1980s.* Retrieved from http://www.mathcurriculumcenter.org/CCM/summaries/agenda4action_summary.pdf

Center for the Study of Mathematics Curriculum. (2005). *Summary of A nation at risk: The imperative for educational reform.* Retrieved from http://www.mathcurriculumcenter.org/PDFS/CCM/summaries/NationAtRisk.pdf

Chodura, S., Kunh, J., & Holling, H. (2015). Interventions for children with mathematical difficulties: A meta-analysis. *Zeitschrift für Psychologie, 223*, 129–144. doi: 10.1027/2151-2604/a000211

*Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Cary, M. S. (2014). Examining the efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Learning Disabilities, 49*, 152–165. doi: 10.1177/0022219414538514

*Clarke, B., Doabler, C., Smolkowski, K., Nelson, E. K., Fien, H., Baker, S. K., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Research on Educational Effectiveness, 9*, 607–634. doi: 10.1080/19345747.2015.1116034

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

Clarke, B., Rolfhus, E., Dimino, J., & Gersten, R. M. (2012). *Assessing student proficiency of number sense (ASPENS).* Longmont, CO: Cambium Learning Group, Sopris Learning.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*, 234–248.

Clarke, B., Smolkowski, K., Baker, S. K., Fien, H., Doabler, C. T., & Chard, D. J. (2011). The impact of a comprehensive tier I core kindergarten program on the achievement of students at risk in mathematics. *The Elementary School Journal, 111*, 561–584. doi: 10.1086/659033

*Clements, D. H. (1984). Training effects on the development and generalization of Piagetian logical operations and knowledge of number. *Journal of Educational Psychology, 76*, 766–776. doi: 10.1037/0022-0663.76.5.766

Clements, D. H., & Sarama, J. (2009). Learning trajectories in early mathematics–sequences of acquisition and teaching. *Encyclopedia of Language and Literacy Development (July 2010*), 1–7.

*Codding, R. S., Chan-Iannetta, L., George, S., Ferreira, K., & Volpe, R. (2011). Early number skills: Examining the effects of class-wide interventions on kindergarten performance. *School Psychology Quarterly*, *26*, 85–96. doi: 10.1037/a0022661

Codding, R. S., Hilt-Panahon, A., Panahon, C., & Benson, J. L. (2009). Addressing mathematics computation problems: A review of simple and moderate intensity interventions. *Education and Treatment of Children, 32*, 279–312. doi: 10.1353/etc.0.0053

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

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

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. Oxford University Press: New York, NY.

DeLoach, D. (2012). Effects of a prekindergarten mathematics intervention on mathematical abilities of preschoolers with low socioeconomic status. (*Unpublished doctoral dissertation*). Walden University, Minneapolis, Minnesota.

Desoete, A., Ceulemans, A., De Weerdt, F., & Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *British Journal of Educational Psychology, 82*, 64–81. doi:10.1348/2044-8279.002002

Desoete, A., & Grégoire, J. (2006). Numerical competence in young children and in children with mathematics learning disabilities. *Learning and Individual Differences, 16*, 351–367. doi:10.1016/j.lindif.2006.12.006

*Desoete, A., & Praet, M. (2013). Inclusive mathematics education: The value of a computerized look-ahead approach in kindergarten. A randomized controlled study. *Transylvanian Journal of Psychology*.

Desoete, A., Stock, P., Schepens, A., Baeyens, D., & Roeyers, H. (2009). Classification, seriation, and counting in grades 1, 2, and 3 as two-year longitudinal predictors for low achieving in numerical facility and arithmetical achievement? *Journal of Psychoeducational Assessment, 27*, 252−264. doi:10.1177/0734282908330588

Doabler, C. T., Clarke, B., & Fien, H. (2012). *ROOTS Assessment of Early Numeracy Skills*. Unpublished measure, University of Oregon, Eugene.

Doabler, C. T., Clarke, B., Fein, H., Baker, S., Kosty, D. B., & Strand Cary, M. (2015). The science behind curriculum development and evaluation: Taking a design science approach in the production of a tier 2 mathematics curriculum. *Learning Disability Quarterly, 38*, 97–111. doi: 10.1177/0731948713520555

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

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., …Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428–1446. doi: 10.1037/0012-1649.43.6.1428

*Dyson, N., Jordan, N. C., Beliakoff, A., & Hassinger-Das, B. (2015). A kindergarten number-sense intervention with contrasting practice conditions for low-achieving children. *Journal for Research in Mathematics Education*, *46*, 331–370. doi: 10.5951/jresematheduc.46.3.0331

*Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities*, *46*, 166–181. doi: 10.1177/0022219411410233

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*, 1502-1505. doi: 10.1126/science.1255484

Friso-van den Bos, I., Kroesbergen, E. H., Van Luit, J. E., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., & Van Lieshout, E. C. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology, 134*, 12–29. doi:10.1016/j.jecp.2015.02.002

Frye, D., Baroody, A. J., Burchinal, M., Carver, S. M., Jordan, N. C., & McDowell, J. (2013). Teaching math to young children. Educator's practice guide. What Works Clearinghouse. NCEE 2014-4005. *National Center for Education Evaluation and Regional Assistance*.

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

Fuchs, L. S., Geary, D., Compton, D. L., Fuchs, D., Hamlett, C. L., & Bryant, J. D. (2010). The contribution of numerosity and domain-general abilities to school readiness. *Child Development, 81*, 1520–1533. doi: 10.1111/j.1467-8624.2010.01489.x

*Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergartners' mathematical development: Effects of peer-assisted learning strategies. *The Elementary School Journal*, *101*, 495–510. doi: 10.1086/499684

Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1990). *Curriculum-based math computation and concepts/applications*. Unpublished measure.

Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003). *Fact fluency assessment.* Unpublished measure.

*Fuchs, L. S., Fuchs, D., Yazdian, L., & Powell, S. R. (2002). Enhancing first-grade children's mathematical development with peer-assisted learning strategies. *School Psychology Review*, *31*, 569–583.

*Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Schatschneider, C., Hamlett, C. L., ... Bryant, J. D. (2013). Effects of first-grade number knowledge tutoring with contrasting forms of practice. *Journal of Educational Psychology*, *105*, 58–77. doi: 10.1037/a0030127

Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, J. C. (1987). *Stanford 7 Plus* (Primary 1). San Antonio, TX: Harcourt, Brace, Jovanovich.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The number sets test. *Journal of Psychoeducational Assessment, 27*, 265–279. doi: 10.1177/0734282908330592

Geary, D. C., Boykin, A. W., Embretson, S., Reyna, V., Siegler, R., Berch, D. B., & Graban, J. (2008). Report of the task group on learning processes. *National mathematics advisory panel, reports of the task groups and subcommittees*, 4-1.

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology, 77,* 236–263. doi:10.1006/jecp.2000.2561

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology, 104*, 206–223. doi: 10.1037/a0025398

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Developmental Neuropsychology, 33,* 277–299. doi:10.1080/87565640801982361

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

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

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

Gersten, R., Rolfhus, E., Clarke, B., Decker, L. E., Wilkins, C., & Dimino, J. (2015). Intervention for first graders with limited number knowledge: Large-scale replication of a randomized controlled trial. *American Educational Research Journal*, *52*, 516–546. doi: 10.3102/0002831214565787

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early numeracy mathematics ability* (3rd ed.). Austin, TX: Pro-Ed.

Grégoire, J., Noel, M., & Van Nieuwenhoven, C. (2004). TEDIMATH Vlaanderen: Test in functie van de diagnostiek van basismathematische kennis en vaardigheden [Functional diagnostic test of basic mathematical knowledge and skills]. *Brussels, Belgium: TEMA*.

Gupta, S. K. (2011). Intention-to-treat: A review. *Perspectives in Clinical Research, 2*, 109 – 112. doi: 10.4103/2229-3485.83221

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2009). Toward developing a science of treatment integrity: Introduction to a special series. *School Psychology Review, 38*, 445–459.

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

*Hansmann, P. (2013). Early numeracy intervention: Does quantity discrimination really work? (*Unpublished doctoral dissertation*). Oklahoma State University, Stillwater, Oklahoma.

*Hassinger-Das, B., Jordan, N. C., & Dyson, N. (2015). Reading stories to learn math: Mathematics vocabulary instruction for children with early numeracy difficulties. *The Elementary School Journal*, *116*, 242–264. doi: 10.1086/683986

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

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analysis. *British Medical Journal*, *327*, 557–560.

Holmes, W., & Dowker, A. (2013). Catch up numeracy: A targeted intervention for children who are low-attaining in mathematics. *Research in Mathematics Education, 15*, 249–265. doi: 10.1080/14794802.2013.803779

Holt, E. W., McGrath, D. J., & Herring, W. L. (2007). *Timing and duration of student participation in special education in the primary grades*. National Center for Education Statistics, Institute of Education Sciences, Department of Education.

Jitendra, A. K., Nelson, G., Pulles, S. M., Kiss, & Houseworth, J. (2016). Is mathematical representation of problems an evidence-based strategy for students with mathematics difficulties? *Exceptional Children, 83*, 8–25. doi: 10.1177/0014402915625062

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

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, *39*, 181–185.

Jordan, N. C., & Hanich, L. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities, 33,* 567–578. doi:10.1177/002221940003300605

Jordan, N. C., & Hanich, L. B. (2003). Characteristics of children with moderate mathematics deficiencies: A longitudinal perspective. *Learning Disabilities Research and Practice, 18*, 213–221. doi:10.1111/1540-5826.00076

Jordan, N. C., Kaplan, D., Nabors Oláh, L., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*, 153–175. doi: 10.1111/j.1467-8624.2006.00862.x

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

Judge, S., & Watson, S. M. (2011). Longitudinal outcomes for mathematics achievement for students with learning disabilities. *The Journal of Educational Research*, *104*, 147–157. doi:10.1080/00220671003636729

Jung, M., Hartman, P., Smith, T., & Wallace, S. (2013). The effectiveness of teaching number relationships in preschool. *International Journal of Instruction, 6*, 167–178.

*Kaufmann, L., Delazer, M., Pohl, R., Semenza, C., & Dowker, A. (2005). Effects of a specific numeracy educational program in kindergarten children: A pilot study. *Educational Research and Evaluation, 5*, 405–431. doi: 10.1080/13803610500110497

Kaufmann, L., Handl, P., & Thöny, B. (2003). Evaluation of a numeracy intervention program focusing on basic numerical knowledge and conceptual knowledge: A pilot study. *Journal of Learning Disabilities, 36*, 564–573. doi: 10.1177/00222194030360060701

Kazdin, A. E. (1977). Artifact, bias, and complexity of assessment: The ABCs of reliability. *Journal of Applied Behavior Analysis*, *10*, 141–150.

Kidd, J. K., Pasnak, R., Gadzichowski, M., Ferral-Like, M., & Gallington, D. (2008). Enhancing early numeracy by promoting the abstract thought involved in the oddity principle, seriation, and conservation. *Journal of Advanced Academics, 19*, 164–200. doi: 10.4219/jaa-2008-780

Klein, A., Starkey, P., Clements, D., Sarama, J., & Iyer, R. (2008). Effects of a pre-kindergarten mathematics intervention: A randomized experiment. *Journal of Research on Educational Effectiveness, 1*, 155–178. doi: 10.1080/19345740802114533

Krajewski, K., & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction*, *19*, 513–526. doi:10.1016/j.learninstruc.2008.10.002

Kroesbergen, E. H., van't Noordende, J. E., & Kolkman, M. E. (2014). Training working memory in kindergarten children: Effects on working memory and early numeracy. *Child Neuropsychology*, *20*, 23–37. doi: 10.1080/09297049.2012.736483

Kroesbergen, E. H., & Van Luit, J. E. H. (2003). Mathematics interventions for children with special education needs: A meta-analysis. *Remedial and Special Education, 24*, 97–114. doi: 10.1177/07419325030240020501

Kunsch, C. A., Jitendra, A. K., & Sood, S. (2007). The effects of peer-mediated instruction in mathematics for students with learning problems: A research synthesis. *Learning Disabilities Research & Practice*, *22*, 1–12. doi: 10.1111/j.1540-5826.2007.00226.x

*Kyttälä, M., Kanerva, K., & Kroesbergen, E. (2015). Training counting skills and working memory in preschool. *Scandinavian Journal of Psychology*, *56*, 363–370. doi: 10.1111/sjop.12221

La Paro, K., Hamre, B., Locasale-Crouch, J., Pianta, R., Bryant, D., Early, D., … Burchinal, M. (2009). Quality in kindergarten classrooms: Observational evidence for the need to increase children's learning opportunities in early education classrooms. *Early Education and Development, 20*, 657–692. doi: 10.1080/10409280802541965

Lee, J. (2012). College for all gaps between desirable and actual P–12 math achievement trajectories for college readiness. *Educational Researcher*, *41*, 43–55. doi: 10.3102/0013189X11432746

Lembke, E. S., & Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and grade 1 students. *Learning Disabilities Research & Practice, 24*, 12–20. doi: 10.1111/j.1540-5826.2008.01273.x

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities, 41*, 451−459. doi: 10.1177/0022219408321126

Lorenz, J. H. (2007). *Hamburger Rechentest.* Manual. Hamburg: Freie und Hansestadt Hamburg, Behörde für Bildung und Sport.

Madden, R., Gardner, E. F., & Collins, C. S. (1987). *Stanford 7 plus* (SESAT 1). San Antonio: Harcourt, Brace, Jovanovich.

Magnuson, K. A., Meyers, M. K., Ruhm, C. J., & Waldfogel, J. (2004). Inequality in preschool education and school readiness. *American Educational Research Journal*, *41*, 115–157. doi: 10.3102/00028312041001115

Malofeeva, E. V. (2005). Meta-analysis of mathematics instruction with young children. (*Unpublished doctoral dissertation*). University of Notre Dame, Indiana.

Malofeeva, E., Day, J., Saco, X., Young, L., & Ciancio, D. (2004). Construction and evaluation of a number sense test with head start children. *Journal of Educational Psychology, 96*, 648–659. doi: 10.1037/0022-0663.96.4.648

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General, 111,* 1–22. doi: 10.1037/0096-3445.111.1.1

Mazzocco, M. M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research & Practice*, *20*, 142–155. doi:10.1111/j.1540-5826.2005.00129.x

Methe, S. A., Hojnoski, R., Clarke, B., Owens, B. B., Lilley, P. K., Politylo, B. C., ... Marcotte, A. M. (2011). Innovations and future directions for early numeracy curriculum-based measurement: Commentary on the special series. *Assessment for Effective Intervention, 36*, 200–209. doi: 10.1177/1534508411414154.

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

*Mohanty, B., & Mishra, S. (1994). Effects of training on the development and generalization of Piagetian number skills and logical operations. *Psychological Studies, 39*, 73–83.

Mononen, R., Aunio, P., Koponen, T., & Aro, M. (2015). A review of early numeracy interventions for children at risk in mathematics. *International Journal of Early Childhood Special Education, 6*, 25–54. doi: 10.20489/intjecse.14355

Monroe, E. E., & Orme, M. P. (2002). Developing mathematical vocabulary. *Preventing School Failure, 46,* 139–142. doi: 10.1080/10459880209603359

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

Mulligan, J., Mitchelmore, M., Kemp, C., Marston, J., & Highfield, K. (2008). Encouraging mathematical thinking through pattern and structure: An intervention in the first year of schooling. *Australian Primary Mathematics Classroom, 13*, 10–15.

Murnane, R. J., Willett, J. B., Braatz, M. J., & Duhaldeborde, Y. (2001). Do different dimensions of male high school students' skills predict labor market success a decade later? Evidence from the NLSY. *Economics of Education Review, 20,* 311–320. Doi: 10.1016/S0272-7757(00)00056-X

Murphy, J., Bates, P., & Anderson, J. (1984). The effect of self-instruction training of counting skills by pre-school handicapped students. *Education and Treatment of Children, 7*, 247–257.

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40*, 458–478. doi: 10.1177/00222194070400050901

National Center for Educational Statistics. (2013a). Interactive graph illustrating the percentage of students scoring within specified ranges on the NAEP. *National Assessment of Educational Progress.* Retrieved from http://nationsreportcard.gov/reading_math_2013/#/student-groups

National Center for Educational Statistics. (2013b). Kindergarten entry status: On-time, delayed-entry, and repeating kindergarteners. Retrieved from https://nces.ed.gov/programs/coe/indicator_tea.asp

National Commission on Excellence in Education, United States Department of Education. (1983). *A nation at risk: The imperative for educational reform. A report to the Nation and the Secretary of Education*. Washington, DC: Author.

National Council of Teachers of Mathematics. (1980). *An agenda for action: Recommendations for school mathematics of the 1980s*. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics* (Vol. 1). Reston, VA: Author.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for pre-kindergarten through grade 8 mathematics: A quest for coherence.* Reston, VA: Author.

National Council of Teachers of Mathematics, Commission on Standards for School Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

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

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity.* Washington, DC: Author.

Nelson, G., & Powell, S. R. (In press). A systematic review of longitudinal studies of mathematics difficulty. *Journal of Learning Disabilities.* Advanced online publication. doi: 10.1177/0022219417714773

*Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction*, *23*, 125–135. doi: 10.1016/j.learninstruc.2012.08.004

Okamato, Y., & Case, R. (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development, 61*, 27–59.

Orwin, R. G., & Cordray, D. S. (1985). Effects of deficient reporting on meta-analysis: a conceptual framework and reanalysis. *Psychological Bulletin, 97*, 134–147. doi: 10.1037/0033-2909.97.1.134

Parsons, S., & Bynner, J. (1997). Numeracy and employment. *Education+Training, 39*, 43–51. doi: 10.1108/00400919710164125

Pasnak, R., Holt, R., Campbell, J. W., & McCutcheon, L. (1991). Cognitive and achievement gains for kindergartners instructed in Piagetian operations. *The*

Journal of Educational Research, *85*, 5–13. doi: 10.1080/00220671.1991.10702807

Pasnak, R., Kidd, J. K., Gadzichowski, M. K., Gallington, D. A., Saracina, R. P., & Addison, K. T. (2009). Promoting early abstraction to promote early literacy and numeracy. *Journal of Applied Developmental Psychology*, *30*, 239–249. doi: 10.1016/j.appdev.2008.12.006

*Passolunghi, M. C., & Costa, H. M. (2016). Working memory and early numeracy training in preschool children. *Child Neuropsychology*, *22*, 81–98. doi: 10.1080/09297049.2014.971726

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

Preschool Curriculum Evaluation Research Consortium (PCER). (2008). Effects of preschool curriculum programs on school readiness (NCER 2008-2009). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Washington, DC: U.S.

Praet, M., & Desoete, A. (2014). Enhancing young children's arithmetic skills through non-intensive, computerised kindergarten interventions: A randomised controlled study. *Teaching and Teacher Education*, *39*, 56–65. doi: 10.1016/j.tate.2013.12.003

PRO-ED. (2007). *Test of Early Mathematics Ability* (3rd ed.). Austin, TX: Author.

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Education Psychology, 105*, 453–464. doi: 10.1037/a0031753

Purpura, D. J., & Lonigan, C. J. (2013). Informal numeracy skills: The structure and relations among numbering, relations, and arithmetic operations in preschool. *American Educational Research Journal, 50*, 178–209. doi: 10.3102/0002831212465332

*Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development*, *79*, 375–394. doi: 10.1111/j.1467-8624.2007.01131.x

Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low-and middle-income preschoolers. *Journal of Applied Developmental Psychology*, *32*, 146–159. doi: 10.1016/j.appdev.2011.02.005

Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development, 24*, 450–472. doi: 10.1016/j.cogdev.2009.09.003

Resnick, L. B. (1989). Developing mathematical knowledge. *American Psychologist*, *44*, 162. doi: 10.1037/0003-066X.44.2.162

Riccomini, P. J., Smith, G. W., Hughes, E. M., & Fries, K. M. (2015). The language of mathematics: The importance of teaching and learning mathematical vocabulary. *Reading and Writing Quarterly, 31,* 235–253. doi: 10.1080/10573569.2015.1030995

*Salminen, J. B., Koponen, T. K., Leskinen, M., Poikkeus, A. M., & Aro, M. T. (2015). Individual variance in responsiveness to early computerized mathematics intervention. *Learning and Individual Differences*, *43*, 124–131. doi: 10.1016/j.lindif.2015.09.002

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*, 662-674. Doi: 10.1016/j.cognition.2008.05.007

*Schacter, J., Shih, J., Allen, C. M., DeVaul, L., Adkins, A. B., Ito, T., & Jo, B. (2016). Math shelf: A randomized trial of a prekindergarten tablet number sense curriculum. *Early Education and Development*, *27*, 74–88. doi: 10.1080/10409289.2015.1057462

*Schopman, E. A., & Van Luit, J. E. (1996). Learning and transfer of preparatory arithmetic strategies among young children with a developmental lag. *Journal of Cognitive Education, 5,* 117–131.

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

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, *11*, 655–661. doi: 10.1111/j.1467-7687.2008.00714.x

*Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, *101*, 545–560. doi: 10.1037/a0014239

*Sood, S., & Jitendra, A. K. (2011). An exploratory study of a number sense program to develop kindergarten students' number proficiency. *Journal of Learning Disabilities, 46*, 328–346. doi: 10.1177/0022219411422380

Spielhagen, F. R. (2006). Closing the achievement gap in math: The long-term effects of eighth-grade algebra. *Journal of Advanced Academics, 18,* 34-59. doi: 10.4219/jaa-2006-344

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science*, *210*, 1033–1035. doi: 10.1126/science.7434014

Starkey, P., Klein, A., & Wakeley, P. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly, 19,* 99 –120. doi:10.1016/j.ecresq.2004.01.002

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

*Toll, S. W. M., & Van Luit, J. E. H. (2012). Early numeracy intervention for low-performing kindergarteners. *Journal of Early Intervention, 34*, 243–264. doi: 10.1177/1053815113477205

*Toll, S. W., & Van Luit, J. E. H. (2013). Accelerating the early numeracy development of kindergartners with limited working memory skills through remedial education. *Research in Developmental Disabilities*, *34*, 745–755. doi: 10.1016/j.ridd.2012.09.003

University of Texas System & Texas Education Agency. (2007). Texas early mathematics inventories-progress monitoring. Austin, TX: Authors.

*Van de Rijt, B. A. M., & Van Luit, J. E. H. (1998). Effectiveness of the additional early mathematics program for teaching children early mathematics. *Instructional Science, 26*, 337–358. doi: 10.1023/A:1003180411209

*Van Luit, J. E. H., & Schopman, E. A. M. (2000). Improving early numeracy of young children with special education needs. *Remedial and Special Education, 21*, 27–40. doi: 10.1177/074193250002100105

Van Luit, J. E. H., Van de Rijt, B. A. M. & Pennings, A. H. (1994). *De Utrechtse Getalbegrip Toets [The Utrecht Early Mathematical Competence Test]*. Doetinchem: Graviant.

Van Luit, J. E. H., & Van de Rijt, B. A. M. (2009). *De Utrechtse Getalbegrip Toets – Revised [The Utrecht Early Mathematical Competence Test – Revised]*. The Netherlands: Graviant.

Van Luit, J. E. H., Van de Rijt, B. A. M., & Aunio, P. (2006) *Early Numeracy Test*, Finnish edition [Lukuk€asitetesti]. Helsinki, Finland: Psykologien kustannus.

Vukovic, R. K. (2012). Mathematics difficulty with and without reading difficulty: Findings and implications from a four-year longitudinal study. *Exceptional Children, 78*, 280–300. doi:10.1177/001440291207800302

Wanzek, J., & Vaughn, S. (2007). Research-based Implications from extensive early reading interventions. *School Psychology Review, 36*, 541–561.

What Works Clearinghouse. (2014). *WWC Procedures and Standards Handbook Version 3.0.* Washington, DC: What Works Clearinghouse, Institute of Education Sciences, US Department of Education; 2014.

Whyte, J. C., & Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Developmental Psychology, 44*, 588–596. doi: 10.1037/0012-1649.44.2.588

Willett, J. B., & Singer, J. D. (1988). Another cautionary note about R 2: Its use in weighted least-squares regression analysis. *The American Statistician*, *42*, 236–238. doi: 10.2307/2685031

*Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain, and Education*, *3*, 224–234. doi: 10.1111/j.1751-228X.2009.01075.x

Witzel, B. S., & Little, M. E. (2016). *Teaching elementary mathematics to struggling learners*. New York, NY: The Guilford Press.

Wynn, K. (1990). Children's understanding of counting. *Cognition, 36*, 155–193. doi: 10.1016/0010-0277(90)90003-3

Xin, Y. P., & Jitendra, A. K. (1999). The effects of instruction in solving mathematical word problems for students with learning problems: A meta-analysis. *The Journal of Special Education*, *32*, 207–225. doi: 10.1177/002246699903200402

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*, B1–B11. doi: 10.1016/S0010-0277(99)00066-9

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants.

Developmental Science, 8, 88–101. doi: 10.1111/j.1467-7687.2005.00395.x

Young-Loveridge, J. M. (2004). Effects on early numeracy of a program using number books and games. *Early Childhood Research Quarterly*, *19*, 82–98. doi: 10.1016/j.ecresq.2004.01.001

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

Appendix A

Coding Procedure for Studies Included in this Meta-Analysis

I. Study Information		
Variable	Code Options	Explanation
Year of publication	Year Also coded for: 0 = published in the last 5 years 1 = published prior to 2012	Year of publication 0 = published during 2012 or later (i.e., publication years 2012, 2013, 2014, 2015, 2016) 1 = published prior to 2012
Country	Country name coded first; for analyses, converted to: 0 = United States 1 = Other	Country where study was conducted; then grouped based on popularity.
Publication type	Selected one: 0 = Peer-reviewed journal 1 = Dissertation 2 = Book/book chapter 3 = Research or technical report 4 = Other	Source of the study; codes are nominal variables (i.e., 0–4).
II. Methodological Characteristics		
Variable	Code Options	Explanation
Study design	Selected one: 0 = Pre-test and post-test 1 = Pre/post-test and delayed post-test 2 = Post-test only	Codes defined as: Pre-test and post-test = Researchers gave both a pre- and post-test to measure the effect of the treatment

	3 = Cross-over design	Pre-/post-test and delayed post-tests = Researchers gave a pre-, post-, and delayed post-test to measure the effect of the intervention Post-test only = Researchers gave only a post-test to measure the effect of the treatment Cross-over = Researchers used a cross-over design; note: only the pre-test and mid-point (cross-over point) were used to measure the effect of the intervention. This was re-coded as pre-test and post-test (0) for analyses
Treatment sample size	Number	Total number of students in the treatment group
Control sample size	Number	Total number of students in the control group
Attrition	Coded as many as possible: a) Total sample attrition b) Treatment attrition Control attrition	Percentage of mortality in total sample, treatment group, and control group; codes are percentages derived from the text or from calculating differences in sample sizes
Assignment to condition	Selected one (assignment, level): 0 = Random, student level 1 = Random, teacher/classroom level 2 = Random, school level 3 = Matching	Type of assignment to the treatment and control condition; codes are nominal variables (i.e., 0–5). For analysis, all random assignment was collapsed into one category.

	4 = Not random, specified 5 = Not reported	
Independent Groups	Selected one: 0 = not independent groups 1 = independent groups	Codes defined as: Not independent = more than one treatment group was compared to the same control group Independent = One treatment group, and only one treatment group, was compared to one control group
Nature of treatment condition	Selected one: 0 = Supplants core 1 = Supplements core 2 = Supplements and supplants 3 = Not reported	Codes defined as: Supplants = the intervention completely replaced the core curriculum (this may include remedial programs for students with disabilities) Supplements = the intervention added to or supplemented the core curriculum and did not replace any part of the core Supplements and supplants = the intervention replaced *part* of the core curriculum (e.g., intervention took place during the last 20 min of regular math instruction) Not reported = no information provided regarding the nature of the treatment condition
Nature of control condition	Selected one: 0 = Business as usual (BAU) controlling time 1 = BAU not controlling for time	Codes defined as: BAU controlling for time = regular classroom mathematics instruction,

	2 = Other math intervention, controlling for time 3 = Active control, controlling for time (non-math instruction or math intervention, but some other activity such as a reading intervention, free time) 4 = Not reported and not able to determine if regular control received regular math instruction, other intervention, etc. For the purpose of analyses, categories were collapsed to BAU (codes 0, 1); active control including math interventions (codes 2, 3) and not reported (code 4).	intervention and control groups received approximately the same total amount of math instruction Business as usual NOT controlling for time = regular mathematics instruction, intervention Other math intervention = treatment and control groups both received math interventions Active control = another activity such as a reading or games controlling for time; Not reported = no information provided by study about control condition
Time of Math Instruction is Controlled	Selected one: 0 = No 1 = Yes 2 = Not able to determine	No = the treatment and control did not receive the same total amount of math instruction time Yes = the treatment and control received the same amount of total math instruction time Not able to determine = not able to determine if the treatment and control received the same amount of math instruction time
Study quality	Reported on a 0-5 scale, with 1 point per item answered, "yes" a) Authors provided thorough information about sample demographics for the treatment and control group *separately*	Codes defined as: a = authors provided relevant information about participant demographics reported separately for treatment and control (at least two of the following): age, grade,

	b) Authors provided relevant demographic information about treatment and control group interventionists c) Authors reported either that fidelity of the treatment condition was collected and how or the fidelity score (measurement) d) Control condition described was described in terms of the mathematics content provided in order to gain insight on the differences between the control and treatment regarding content. e) Attrition For the purpose of analyses, studies were awarded a sum of points between 0 and 5 points and then categorized as: High Quality = 4 or 5 points Medium Quality = 2 or 3 points Low Quality = 0 or 1 points	race/ethnicity, English learner status, disability status, gender, SES status b = authors provided relevant information (at least 2 of the following) about the interventionists: gender, training, teaching credentials, level of education, prior experience; if the intervention was a CAI and the intervention agent was technically a computer program in order to receive a "1" in this category the study must have provided relevant demographic information regarding the supervising teachers of the CAI program c = authors reported that fidelity of the intervention was recorded; if the intervention was CAI fidelity of the intervention is assumed because it is administered via computer program, but in order to receive a score of "1" studies that utilized CAI needed to report attendance or average number of sessions attended by participants. d = authors provided a description of the control condition that either allowed for replication (i.e., reader is able to determine which early numeracy features were present in the control condition). At a minimum, enough information needed to be provided in text that allowed readers to determine whether the control condition mathematics content represented any of

		the three early numeracy domains (number, relations, operations) e = the study provided attrition information for the total sample or separately for groups
III. Participant Characteristics		
Variable	Code Options	Explanation
Age	Mean age of participants	Coded as years; converted "months" to years and "years, months" to years
Grade	Grade	Codes as preschool, kindergarten, and first grade *Note*. Foreign studies that stated participants were in their first of two years of kindergarten were coded as "preschool" and foreign studies that stated participants were in their second of two years of kindergarten were coded as "kindergarten"
Gender	Number, percent of participants identified as: a) Female b) Male	Total number and percent of participants identified in each category.
Race/ethnicity	Number, percent of participants identified as: a) White b) Black/African American c) Asian d) Hispanic/Latino e) Other	Total number and percent of participants identified in each category.

English Learners	Number, percent	Total number and percent of participants identified as English Learners (US studies only).
Free/reduced lunch	Number, percent	Total number and percent of participants identified as receiving free or reduced price lunch (US studies only). For foreign studies, SES information was recorded anecdotally.
Disability	Number, percent	Total number and percent of participants identified as having a documented disability or as receiving special education services.
At-risk type	Selected one: 0 = High-achieving (HA) 1 = Typically achieving (TA) 2 = At-risk, low-income (AR) 4 = math difficulty; less restrictive (MD2) 5 = math difficulty; restrictive (MD1) 6= Documented disability (DIS) For the purpose of analyses, original codes were re-coded as: 0 = Low academic risk (codes of HA, TA) 1 = At-risk due to low SES (codes of AR) 2 = At-risk for MD by performance above the 26th percentile or disability status without indication of performance below the 26th percentile (codes MD2 and DIS-when applicable)	Codes defined by how researchers reported results for subgroups: HA = sample was identified as high achieving by percentile, teacher report, or other method. TA = sample was either not screened for MD and results are reported for the sample not disaggregated by risk, or the sample was identified as typically achieving by percentile, teacher report, or other method. AR = sample was identified as at-risk for due to high percentage of students low-income or qualifying for FRL. MD2= sample identified as MD by performance below a certain percentile

	3 = At-risk for MD by performance at or below the 25th percentile, including students with disability who also had performance in this area (codes MD1 and DIS-when applicable)	(but above the 26th percentile), or alternate modes of identification including percent correct on a screening measure (but no percentile information), and teacher referral. MD1 = sample identified as MD with performance at or below the 25th percentile on a math screening measure (includes cut-off scores reported as *SD*s, converted to approximate percentiles) DIS = sample identified or referred for special education (e.g., students who attend special education schools, authors identify students as having a disability, authors provide disability classification criteria). This code may also include students with disabilities who also scored below a specified criterion on a screening measure
IV. Intervention Characteristics		
Variable	Code Options	Explanation
instructional arrangement	Selected as many codes that apply: 0 = One-on-one instruction 1 = Small group 2 = Mixed, flexible grouping 3 = PALS (specific program) 4 = Not reported	Codes defined as: One-on-one instruction = students received instruction individually, including computer-administered interventions Small group = students received instruction in a small group setting of 2 to 5 students

		Mixed = students received instruction in flexible grouping (changed depending on lesson), including whole class, small groups, and individual PALS = students received instruction in pairs and some whole class instruction, pairs were intentionally created using PALS guidelines Not reported = study did not report instructional arrangement
Intervention agent	Selected one: 0 = Researcher 1 = Teacher 2 = Other school staff 3 = Mixed 4 = Computer-administered 5 = Not reported	Codes defined as: Researcher = a trained researcher or assistant for the specific purpose of conducting the study provided the intervention Teacher = Regular classroom teacher implemented the intervention Other school staff = Other school staff (not specified if intervention agent was the classroom teacher) implemented the intervention material; include staff hired by school to provide interventions such as retired school teachers; includes paraprofessionals Mix = mix of researchers, teachers, and school staff implemented the intervention Computer-administered = intervention content was delivered via computer or electronic device

		Not reported = not reported or not able to determine
Duration	Days, weeks, or months; recoded as weeks. For the purpose of analyses, duration was re-coded as: 0 = duration is 8 weeks or less 1 = duration is greater than 8 weeks	Coded as total duration of the intervention as the time from the first session to the last session (unless specified in the study that total intervention time was specifically a different amount of time)
Intensity	Number of sessions per day, week, month	Coded as the frequency of the intervention (e.g., 2 sessions per week)
Number of sessions	Total number	Coded as the total number of sessions that the intervention lasted; for analyses, this variable was calculated if the total number of sessions was not provided (i.e., total number of weeks × number of sessions per week).
Length of session	Minutes	Coded as the average number of min each intervention session lasted; for analyses, if the session length was provided as a range (e.g., sessions lasted 15 to 20 min, the mean of the two numbers was used).
Total instructional time	Hours, Minutes	Coded as the total number of hours, minutes of instruction (e.g., number of weeks × number of sessions per week × minutes per session). This was a calculated variable for analyses as it was very rarely provided in the study.

		For analysis, studies were given the following code: 0 = 0-9 hours 1 = 10 – 19 hours 2 = 20 or more hours
Instructional features	Selected as many codes that apply: a) Explicit and/or systematic instruction b) Scripted lessons c) Modeling d) Guided practice e) Independent practice f) Pictorial representations g) Concrete representations h) CRA framework i) Corrective feedback j) Reinforcement k) Behavior management plan l) Worksheets m) Game-based n) Book-based o) Included games and/or books Progress monitoring of the treatment and/or control condition	See *Definition of Key Terms* (Chapter 1)
V. Mathematics Content		
Variable	Code Options	Explanation
Numeracy domains	Based on the specific numeracy skills coded (below) as many of the following numeracy domains were selected: a) Number b) Relations	See *Definition of Key Terms* (Chapter 1). If at least one skills in any domain was coded, the intervention was coded as addressing the corresponding domain.

	c) Operations	
Mathematical language	Selected one: 0 = No 1 = Yes	Code defined as: No = no mention of vocabulary or language being explicitly taught Yes = it was clear from text, tables, or example lesson plans that mathematical language or vocabulary was an explicit component of the intervention; explicit refers to that mathematical language or vocabulary was intentionally taught and not simply incorporated into lessons (e.g., including the words "add" or "more" in a lesson plan did not indicate vocabulary was explicitly taught).
Number skills	Selected as many codes that apply: a) Counting sequence (b) One-to-one counting correspondence c) Cardinality d) Counting error identification e) Numeral identification f) Subitizing g) Other	See *Definition of Key Terms* (Chapter 1) Also converted to the number of Number skills in each study.
Relations skills	Selected as many as applicable: a) Matching quantities to numerals, quantities to number words, or number words to numerals b) Numeral comparison c) Set comparison d) Missing number	See *Definition of Key Terms* (Chapter 1) Also converted to the number of Number skills in each study.

	e) Number line estimation f) Number line sequence g) Ordinal numbers h) Other	
Operations skills	Selected as many as applicable: a) Composition/decomposition b) Equivalence c) Simple addition/subtraction with objects d) Simple addition/subtraction without objects (includes number combinations) e) Properties of addition/subtraction f) Place value, base-10 system g) Other	See *Definition of Key Terms* (Chapter 1) Also converted to the number of Number skills in each study.
VI. Dependent Measures		
Variable	Code Options	Explanation
Mathematics measures	Recorded the name and citations of all math measures administered.	Coded the specific name of the measures and the citation as a step toward selecting the primary measure.
Description	Skills measured	Described the measure and what skills were assessed.
Alignment of measures to the intervention (also same code as "type of primary measure")	Selected one for all measures: a) Proximal, comprehensive b) Distal, comprehensive c) Proximal, narrow Distal, narrow	Codes defined as: Proximal = aligned to the intervention (e.g., researcher developed measure, measure of numeracy skills specific to the intervention)

		Distal = not closely aligned to the intervention (e.g., achievement test, calculation) Comprehensive = measures more than one numeracy skill Narrow = measures one specific skills (e.g., numeral identification)
Primary measure	Name of measure	Coded the specific name of the primary measure. Primary measures were selected in order of the alignment to the intervention (e.g., proximal, comprehensive measures were selected over distal, comprehensive, and so on).
Primary measure norm-referenced	Selected one: 0 = Norm-referenced 1 = Not norm-referenced 2 = Not able to determine	Codes defined as: Norm-referenced = the primary measure is norm-referenced and this was determined by author report, or correspondence with the test publisher or author of the measure. Not norm-referenced = the measure is not norm-referenced (e.g., research developed for the purpose of the intervention). This includes measures that are widely available through publishers, but are not norm-referenced. Not able to determine = not enough information to determine whether the measure was norm-referenced.

Progress monitoring in treatment and/or control	Selected one: a) Yes No	Coded yes or no if authors monitored progress on the effect of the intervention (i.e., CBM) during the intervention period. This included administering assessments during the intervention period that were not the pre-test, post-test, or delayed post-test administrations.
Reliability	Reliability estimate	Coded estimates of reliability for all measures; this included: test-retest, split-half, alternate form, and Cronbach's alpha.
Validity	Validity estimates with other measures	Coded estimates of validity of the primary measure with other mathematics measures (i.e., correlations).
Delayed post-test	Selected one: b) Yes c) No	Coded yes or no if delayed post-test was given with the primary measure.
Delayed post-test time	Weeks, months	Recorded the length of time (weeks, months) between the end of the intervention and the delayed post-test
Other measures	List	Listed of all other measures administered (e.g., working memory)
VII. Confidence Codes		
Variable	Code Options	Explanation

General confidence code	Selected one: 0 = Very confident 1 = Mostly confident 2 = Not confident	The purpose of this code was not to code for the quantity of information provided; for example, I did not code for information left out of the study (no race/ethnicity reported), but instead I coded for how confident I am in the codes that I recorded in the coding sheet. 0 = 0 or 1 discrepancies 1 = 2 to 4 discrepancies 2 = 5 or more discrepancies
1) General confidence code discrepancies	Description	Specific discrepancies were recorded.
2) Effect size confidence	Selected one: 2 = No estimation 1 = Some estimation 0 = Highly estimated	The purpose of this code was to record how confident I was in the effect size assigned to each group based on the information reported in the study. 2 = all conventional statistical data are present to calculate effect sizes (*M*, *SD*s) 1 = unconventional statistics reported that require conversion, use of significance tests that require conversion, sample size is not reported appropriate for groups 0 = information available to calculate effect sizes is limited (*N* and *p*-value).

3) Effect size confidence description	Description	I recorded the statistical information provided in the report that was used to calculate the effect size for each group.

Appendix B

Table B1

Summary of Selected and Non-Selected Treatment Groups for the Sensitivity Analysis

Study (year)	Treatment Groups	Rationale for Selection of Treatment Group 1
Desoete & Praet	1) Counting 2) Comparison	Although counting and comparison skills are both critical to children's foundational understanding of numeracy content, the Counting treatment condition also introduced participants to very simple addition and subtraction with objects; therefore, this intervention group was determined to be representative of more early numeracy domains and skills.
Dyson et al. (2015)	1) Number List 2) Fact Practice	The two treatment groups were nearly identical other than the final 5 minutes of each intervention session. The Number List treatment group was selected because this treatment group practiced strategies for counting on, writing numerals, and matching numerals to quantities. The Fact Practice treatment group focused on fluency. In order to align the selection of treatment groups with the original intent studies that met inclusion in this study, the treatment condition with the fluency component was not selected.
Fuchs et al. (2013)	1) Non-speeded Practice 2) Speeded Practice	The two treatment groups were nearly identical other than the final 5 minutes of each intervention session. The Non-Speeded Practice treatment group was selected because for the final 5 minutes of the intervention period, this treatment group worked on activities to reinforce concepts and apply skills learned in the day's lesson as they were related to other concepts and principles. In contract, the Speeded Practice condition practiced fact fluency through drill flashcards. In order to align the selection of treatment groups with the original intent studies that met inclusion in this study, the treatment condition with the fluency component was not selected.

Table B1 continued…

Study (year)	Treatment Groups	Rationale for Selection of Treatment Group 1
Hansmann (2013)	1) Traditional QD 2) Triangle QD	The Traditional QD treatment group was selected because the presentation of the quantity discrimination task represents how quantity discrimination is typically assessed in early numeracy contexts, this approach was used by the majority of other studies that also taught quantity discrimination.
Hassinger-Das et al. (2015)	1) Traditional Numeracy 2) Storybook Numeracy	The Traditional Numeracy treatment group was selected because it was more closely aligned to how numeracy content was delivered in the majority of studies, while the Storybook Numeracy treatment group had a focus on teaching numeracy concepts through books.
Obersteiner et al. (2013)	1) Both 2) Exact 3) Approximate	The Exact treatment group focused on developing exact representations of organized dot patterns and the Approximate treatment group focused on developing approximate analogue magnitudes. The Both treatment group was a mix of Exact and Approximate approaches. The Both treatment group was selected because this condition required participants to activate exact and approximate number processing and the Both condition represents a blend of the two main approaches to enhancing number line representation.

Note. The “selected” treatment group is noted as “1)” for each study.

Appendix C
Demographics

Table C1
Demographic Information for Studies that Reported Demographics Separately for Treatment and Control Groups

	N		Males		ELs		FRL		White		Black		Hispanic[a]	
Study (year)	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn
Bryant et al. (2011)	139	65	44%	55%	5%	9%	50%	52%	37%	32%	27%	22%	33%	40%
Clarke, Doabler, Smolkowski, Baker, et al. (2016)	63	63	51%	55%	72%	56%	—	—	—	—	—	—	—	—
Clarke, Doabler, Smolkowski, Kurtz Nelson, et al. (2016)	203	87	45%	41%	32%	31%	—	—	56%	60%	4%	5%	33%	32%
Clarke et al. (2014)	44	45	48%	64%	14%	22%	71%	69%	91%	80%			14%	27%
Codding et al. (2011)	30	35	—	—	—	—	—	—	—	—	—	—	—	—
Desoete & Praet (2013) [*TA*]	24	16	—	—	NA	NA	NA	NA	—	—	—	—	—	—
Desoete & Praet (2013) [*MD*]	62	33	—	—	NA	NA	NA	NA	—	—	—	—	—	—

	N		Males		ELs		FRL		White		Black		Hispanic[a]	
Study (year)	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn
Doabler et al. (2016)	229	90	47%	58%	23%	28%	—	—	91%	84%	6%	7%	49%	51%
Dyson et al. (2015) [*number list*][b]	40	42	50%	52%	43%	43%	78%	86%	25%	19%	23%	24%	48%	52%
Dyson et al. (2015) [*fact practice*][b]	44	42	45%	52%	39%	43%	82%	86%	20%	19%	27%	24%	52%	52%
Dyson et al. (2013)	56	65	61%	54%	25%	25%	—	—	7%	5%	57%	55%	36%	38%
Fuchs et al. (2005)	64	63	47%	52%	—	—	56%	51%	45%	44%	50%	49%	5%	6%
Fuchs et al. (2001) [*HA*]	14	17	50%	58%		6%	7%	24%	86%	65%	14%	29%		
Fuchs et al. (2001) [*TA*]	49	52	47%	48%			47%	54%	24%	21%	4%	10%		
Fuchs et al. (2001) [*MD*]	13	8	69%	63%	8%		62%	38%	15%		77%	100%		
Fuchs et al. (2001) [*DIS*]	8	7	63%	71%	13%	14%	63%	43%	38%		63%	63%		
Fuchs et al. (2002) [*HA*]	47	46	60%	48%	2%	4%	34%	41%	45%	54%	51%	37%	4%	9%
Fuchs et al. (2002) [*TA*]	75	75	55%	57%	75%	78%	61%	53%	32%	29%	43%	48%	27%	23%

	N		Males		ELs		FRL		White		Black		Hispanic[a]	
Study (year)	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn
Fuchs et al. (2002) [*MD*]	39	43	59%	58%	38%	49%	72%	70%	23%	16%	38%	37%	38%	47%
Fuchs et al. (2013) [*non-speeded practice*][b]	190	206	47%	50%	—	—	79%	87%	21%	17%	67%	73%	8%	6%
Fuchs et al. (2013) [*speeded practice*][b]	195	206	51%	50%	—	—	85%	87%	23%	17%	68%	73%	7%	6%
Jordan et al. (2012)	42	44	60%	57%	29%	18%	—	—	14%	14%	45%	43%	40%	41%
Kaufmann et al. (2005)	17	17	47%	53%	NA	NA	NA	NA	—	—	—	—	—	—
Passolunghi & Costa (2016)	15	18	60%	50%	NA	NA	NA	NA	—	—	—	—	—	—
Ramani & Siegler (2008)	68	56	44%	48%	—	—	PHS	PHS	43%	39%	50%	55%		
Salminen et al. (2015)	13	8	31%	50%	NA	NA	NA	NA	—	—	—	—	—	—
Schacter et al. (2016)	45	41	46%	50%	—	—	PHS	PHS	10%	8%	42%	40%	46%	50%
Schopman & Van Luit (1996)	40	20	—	—	NA	NA	NA	NA	—	—	—	—	—	—

Study (year)	*N*		Males		ELs		FRL		White		Black		Hispanic[a]	
	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn
Siegler & Ramani (2009) [*linear*][b]	30	29	40%	45%	—	—	PHS	PHS	16	20	40%	31%		
Sood & Jitendra (2011) [*TA*]	36	22	47%	50%	—	—	53%	45%	42%	50%	44%	41%	8%	9%
Sood & Jitendra (2011) [*MD*]	25	18	68%	56%	—	—	80%	56%	16%	39%	44%	28%	36%	33%
Sood & Jitendra (2011) [*MD*]	25	18	68%	56%	—	—	80%	56%	16%	39%	44%	28%	36%	33%
Toll & Van Luit (2012) [*MD 0-24th*]	25	21	52%	29%	NA	NA	NA	NA	—	—	—	—	—	—
Toll & Van Luit (2012) [*MD 25-50th*]	27	25	48%	32%	NA	NA	NA	NA	—	—	—	—	—	—
Toll & Van Luit (2013) [*MD*]	115	87	53%	49%	NA	NA	NA	NA	—	—	—	—	—	—
Toll & Van Luit (2013) [*MD-WM*]	31	43	45%	72%	NA	NA	NA	NA	—	—	—	—	—	—

	N		Males		ELs		FRL		White		Black		Hispanic[a]	
Study (year)	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn	Tr	Cn
Van de Rijt & Van Luit (1998) [*guiding*]	27	29	—	—	—	—	—	—	—	—	—	—	—	—
Van de Rijt & Van Luit (1998) [*structure*]	26	24	—	—	—	—	—	—	—	—	—	—	—	—
Van Luit & Schopman (2000)	62	62	65%	66%	NA	NA	NA	NA	87%	85%				
Wilson et al. (2009)	26	26	—	—	—	—	—	—	—	—	—	—	—	—

Note. Different treatment groups and risk groups (e.g., MD, TA, HA) are specified in italics with the study (year) information. TA = typically achieving, MD = mathematics difficulty, HA = high achieving; DIS = students with disabilities. Regarding percentages, not all columns within categories add to 100% due to rounding. CN = Control group; ELs = percentage of English Learners; FRL = percentage of free and/or reduced price lunch; NA = not applicable (e.g., ELs and FRL for foreign studies); PHS = participants drawn from Head Start Centers; Tr = Treatment group. Empty columns reflect values of "zero" while columns with "—" reflect information that was not reported for disaggregated groups.

[a] Some studies reported students race and ethnicity (Hispanic) as exclusive (i.e., students could only be identified as White or Hispanic), while other studies identified students as White and Hispanic. Therefore, percentages for race and ethnicity may not add to 100%.

[b] Studies marked with a "b" indicate that the same control group was used for both treatment groups.

Table C2

Demographic Information for Studies that Reported Demographics for the Full Sample

Study (year)	*N*	Male	ELs	FRL	White	Black	Hispanic
Clements (1984)	45	56%	—	—	—	—	—
Hansmann (2013)	123	47%	—	41%	78%	7%	8%
Hassinger-Das et al. (2015)	124	52%	55%	83%	18%	18%	63%
Kyttälä et al. (2015)	61	52%	NA	NA	—	—	—
Mohanty & Mishra (1994)	30	—	NA	NA	—	—	—
Obersteiner et al. (2013)	147	49%	NA	NA	—	—	—

Note. Regarding percentages, not all columns within categories add to 100% due to rounding. ELs = percentage of English Learners; FRL = percentage of free and/or reduced price lunch; NA = not applicable (e.g., ELs and FRL for foreign studies). Empty columns reflect values of "zero" while columns with "—" reflect information that was not reported for disaggregated groups.

Appendix D

Table D1

Instructional Features of Interventions

Study (year)	Explicit instruction	Modeling	Guided practice	Ind. practice[a]	Corrective feedback	Reinforce-ment	Script[b]
Bryant et al. (2011)	X	X	X	X	X	X	X
Clarke, Doabler, Smolkowski, Baker, et al. (2016)	X	X	X	X	X		X
Clarke, Doabler, Smolkowski, Kurtz Nelson, et al. (2016)	X	X	X	X	X		X
Clarke et al. (2014)	X	X	X	X	X		X
Clements (1984)							
Codding et al. (2011)	X				X	X	X
Desoete & Praet (2013) [*counting*]	X			X	X	X	X
Desoete & Praet (2013) [*compare*]	X			X	X	X	X
Doabler et al. (2016)	X	X	X	X	X		X
Dyson et al. (2015) [*number list*]	X	X	X	X			X
Dyson et al. (2015) [*fact retrieval*]	X	X	X	X			X
Dyson et al. (2013)	X				X		X
Fuchs et al. (2005)	X			X		X	X
Fuchs et al. (2001)	X				X	X	X
Fuchs et al. (2002)	X	X		X	X	X	
Fuchs et al. (2013) [*speeded*]	X	X	X	X	X		X
Fuchs et al. (2013) [*non-speeded*]	X	X	X	X	X		X
Hansmann (2013) [*traditional QD*]	X			X			X

Study (year)	Explicit instruction	Modeling	Guided practice	Ind. practice[a]	Corrective feedback	Reinforce-ment	Script[b]
Hansmann (2013) [*triangle QD*]	X			X			X
Hassinger-Das et al. (2015) [*numeracy*]	X				X		X
Hassinger-Das et al. (2105) [*story problems*]	X		X				X
Jordan et al. (2012)	X				X		X
Kaufmann et al. (2005)	X	X		X		X	
Kyttälä et al. (2015)							
Mohanty & Mishra (1994)	X			X			
Obersteiner et al. (2013) [*exact*]	X		X	X		X	X
Obersteiner et al. (2013) [*approximate*]	X		X	X		X	X
Obersteiner et al. (2013) [*both*]	X		X	X		X	X
Passolunghi & Costa (2016)							
Ramani & Siegler (2008)	X	X	X		X		
Salminen et al. (2015)	X				X	X	X
Schacter et al. (2016)	X				X	X	X
Schopman & Van Luit (1996)	X						
Siegler & Ramani (2009)	X	X	X		X		
Sood & Jitendra (2011)	X	X	X	X	X		X
Toll & Van Luit (2012)	X	X	X	X	X		X
Toll & Van Luit (2013)	X	X	X				X

Study (year)	Explicit instruction	Modeling	Guided practice	Ind. practice[a]	Corrective feedback	Reinforce-ment	Script[b]
Van de Rijt & Van Luit (1998) [*structured*]	X			X			
[*guiding*]	X						
Van Luit & Schopman (2000)	X	X	X	X			
Wilson et al. (2009)	X		X	X		X	X

Note. QD = quantity discrimination.
[a]Independent practice; [b]Scripted lessons.

Table D1 Continued…

Study (year)	CRA[c]	Concrete manip.[d]	Visual rep.[e]	Behavior plan	Game or book-based	Included games/books
Bryant et al. (2011)	X	X	X	X		
Clarke, Doabler, Smolkowski, Baker, et al. (2016)		X	X			
Clarke, Doabler, Smolkowski, Kurtz Nelson, et al. (2016)		X	X			
Clarke et al. (2014)	X	X	X			
Clements (1984)						
Codding et al. (2011)		X	X			
Desoete & Praet (2013) [*counting*]					X	
Desoete & Praet (2013) [*compare*]					X	
Doabler et al. (2016)			X			

Study (year)	CRA[c]	Concrete manip.[d]	Visual rep.[e]	Behavior plan	Game or book-based	Included games/books
Dyson et al. (2015) [*sequence*]		X	X			
Dyson et al. (2015) [*fact retrieval*]		X	X			
Dyson et al. (2013)		X	X			X
Fuchs et al. (2005)	X	X	X	X		X
Fuchs et al. (2001)		X	X			
Fuchs et al. (2002)		X	X			X
Fuchs et al. (2013) [*speeded*]		X	X	X		X
Fuchs et al. (2013) [*non-speeded*]		X	X	X		X
Hansmann (2013) [*traditional QD*]						
Hansmann (2013) [*triangle QD*]						
Hassinger-Das et al. (2015) [*numeracy*]		X	X			X
Hassinger-Das et al. (2015) [*story problems*]		X	X		X	X
Jordan et al. (2012)		X	X			X
Kaufmann et al. (2005)	X	X	X		X	
Kyttälä et al. (2015)			X			X
Mohanty & Mishra (1994)		X	X			
Obersteiner et al. (2013) [*exact*]			X		X	
Obersteiner et al. (2013) [*approximate*]			X		X	

Study (year)	CRA[c]	Concrete manip.[d]	Visual rep.[e]	Behavior plan	Game or book-based	Included games/books
Obersteiner et al. (2013) [*both*]			X		X	
Passolunghi & Costa (2016)		X	X		X	
Ramani & Siegler (2008)			X		X	
Salminen et al. (2015)	X	X	X		X	
Schacter et al. (2016)		X	X		X	
Schopman & Van Luit (1996)		X			X	
Siegler & Ramani (2009)			X		X	
Sood & Jitendra (2011)	X	X	X			X
Toll & Van Luit (2012)	X	X	X			X
Toll & Van Luit (2013)	X	X	X			X
Van de Rijt & Van Luit (1998) [*guiding*]	X	X	X			
Van de Rijt & Van Luit (1998) [*structured*]						
Van Luit & Schopman (2000)	X	X	X			
Wilson et al. (2009)			X		X	

Note. QD = quantity discrimination.
[c]CRA = concrete-representations-abstract framework; [d]Concrete manipulatives; [e]Visual representations.

Appendix E

Table E1

Dependent Measure Characteristics and Other Measures

	Primary Mathematics Outcome Measure				Other Measures	
Study (year)	Broad proximal	Broad distal	Narrow proximal	Narrow distal	More than 1 math measure	Non-math measures
Bryant et al. (2011)	X				X	
Clarke, Doabler, Smolkowski, Baker, et al. (2016)		X			X	
Clarke, Doabler, Smolkowski, Kurtz Nelson, et al. (2016)	X				X	
Clarke et al. (2014)	X				X	
Clements (1984)	X					X
Codding et al.(2011)		X			X	
Desoete & Praet (2013) [*counting*; TA]				X	X	
Desoete & Praet (2013) [*counting*; MD]				X		
Desoete & Praet (2013) [*comparison*; TA]				X		
Desoete & Praet (2013) [*comparison*; MD]				X		
Doabler et al. (2016)	X				X	
Dyson et al. (2015) [*number sequence*]	X				X	X
Dyson et al. (2015) [*fact retrieval*]	X				X	X
Dyson et al. (2013)	X				X	X
Fuchs et al. (2005)		X			X	
Fuchs et al. (2001) [HA]		X			X	
Fuchs et al. (2001) [TA]		X			X	

Study (year)	Primary Mathematics Outcome Measure				Other Measures	
	Broad proximal	Broad distal	Narrow proximal	Narrow distal	More than 1 math measure	Non-math measures
Fuchs et al. (2001) [MD]		X			X	
Fuchs et al. (2001) [DIS]		X			X	
Fuchs et al. (2002) [HA]	X				X	
Fuchs et al. (2002) [TA]	X				X	
Fuchs et al. (2002) [MD]	X				X	
Fuchs et al. (2013) [*speeded practice*]		X			X	X
Fuchs et al. (2013) [*non-speeded practice*]		X			X	X
Hansmann (2013) [*traditional* QD]		X			X	
Hansmann (2013) [*triangle* QD]		X			X	
Hassinger-Das et al. (2015) [*numeracy*]	X				X	X
Hassinger-Das et al. (2015) [*story problems*]		X			X	X
Jordan et al. (2012)	X				X	X
Kaufmann et al. (2005)	X				X	
Kyttälä et al. (2015)			X			X
Mohanty & Mishra (1994)	X					X
Obersteiner et al. (2013) [*exact*]		X			X	
Obersteiner et al. (2013) [*approximate*]		X			X	
Obersteiner et al. (2013) [*both*]		X			X	
Passolunghi & Costa (2016)	X					X
Ramani & Siegler (2008)			X		X	
Salminen et al. (2015)			X		X	

Study (year)	Primary Mathematics Outcome Measure				Other Measures	
	Broad proximal	Broad distal	Narrow proximal	Narrow distal	More than 1 math measure	Non-math measures
Schacter et al. (2016)	X					
Schopman & Van Luit (1996)	X					X
Siegler & Ramani (2009)			X		X	
Sood & Jitendra (2011) [TA]	X				X	
Sood & Jitendra (2011) [MD]	X				X	
Toll & Van Luit (2012) [MD, <25th pc.]	X					X
Toll & Van Luit (2012) [MD, 26-50th pc.]	X					X
Toll & Van Luit (2013) [MD]	X					X
Toll & Van Luit (2013) [MDWM]	X					X
Van de Rijt & Van Luit (1998) [*guiding*]	X					
Van de Rijt & Van Luit (1998) [*structured*]	X					
Van Luit & Schopman (2000)	X				X	X
Wilson et al. (2009)			X		X	

Note. Different treatment groups are specified in italics and risk groups are also specified (e.g., MD, TA, HA) with the study (year) information. DIS = students with disabilities; HA = high achieving; MD = mathematics difficulty; MDWM = mathematics difficulty and low working memory scores; pc. = percentile; QD = quantity discrimination; TA = typically achieving.

Appendix F

Table F1

Simple Comparisons of Effects by Dependent Measure

Characteristic				95% Confidence Interval		Within Group		Between Groups		
Full Sample	*n*	mean (*g*)	*SE*	Lower	Upper	*Q*	*p*(*Q*)	*Q*	*df*	*p*(*Q*)
Primary Measure								21.85	3	0.00
Proximal, Broad	27	0.84	0.08	0.68	1.00	130.30	0.00			
Distal, Broad	15	0.29	0.11	0.09	0.50	9.10	0.83			
Proximal, Narrow	5	0.56	0.20	0.17	0.95	7.53	0.11			
Distal, Narrow	4	1.16	0.24	0.70	1.63	6.15	0.11			
Norm-Referenced								6.91	2	0.03
No	28	0.84	0.09	0.66	1.02	152.50	0.00			
Yes	20	0.50	0.11	0.29	0.72	61.06	0.00			
Not Reported	3	0.37	0.27	-0.16	0.91	0.67	0.72			
Outliers Removed										
Primary Measure										
Proximal, Broad	25	0.76	0.07	0.64	0.89	73.94	0.00	26.60	3	0.00
Distal, Broad	15	0.30	0.08	0.13	0.46	9.10	0.83			
Proximal, Narrow	5	0.58	0.17	0.25	0.90	7.53	0.11			
Distal, Narrow	4	1.15	0.20	0.75	1.55	6.15	0.11			
Norm-referenced								5.12	2	0.08
No	26	0.75	0.08	0.59	0.91	92.96	0.00			
Yes	20	0.51	0.09	0.32	0.69	61.06	0.00			
Not Reported	3	0.37	0.24	-0.09	0.84	0.67	0.72			

Note. Full sample (*N* = 51 effect sizes); outliers removed sample (*N* = 49 effect sizes). Within Group *Q* test represents heterogeneity within the variable noted (e.g., Proximal, Broad measure); Across Groups *Q* test represents heterogeneity across variable coding of (i.e., the comparison between all primary measure outcome types).

Appendix G

Table G1

Simple Comparisons of Effects by Numeracy Domain

Early Numeracy Domain				95% CI			
Full Sample	*n*	Mean *g*	*SE*	Lower	Upper	*Q*	*p*(*Q*)
Number Domain							
No	7	0.38	0.18	0.02	0.73	8.46	0.21
Yes	44	0.72	0.07	0.58	0.87	204.16	0.00
0 or 1 skills	13	0.35	0.12	0.12	0.59	15.98	0.19
2 or 3 skills	20	0.97	0.10	0.77	1.17	71.50	0.00
4 or more skills	18	0.58	0.10	0.39	0.78	80.00	0.00
Relations Domain							
No	2	1.54	0.37	0.81	2.26	0.26	0.61
Yes	49	0.65	0.07	0.51	0.78	205.2	0.00
0 or 1 skills	10	0.80	0.15	0.51	1.10	65.48	0.00
2 or 3 skills	30	0.59	0.09	0.42	0.76	101.07	0.00
4 or more skills	11	0.78	0.14	0.51	1.04	27.05	0.00
Operations Domain							
No	12	0.70	0.14	0.42	0.98	26.04	0.01
Yes	39	0.67	0.08	0.52	0.82	194.61	0.00
0 or 1 skills	25	0.84	0.10	0.65	1.04	126.43	0.00
2 or 3 skills	16	0.49	0.12	0.25	0.73	36.23	0.00
4 or more skills	10	0.60	0.14	0.33	0.87	44.13	0.00
Mathematical Language							
No	41	0.64	0.08	0.49	0.79	177.89	0.00
Yes	10	0.81	0.15	0.52	1.10	30.97	0.00

Table G1 Continued…

Early Numeracy Domain				95% CI			
Outliers Removed	*n*	Mean *g*	*SE*	Lower	Upper	*Q*	*p*(*Q*)
Number Domain							
No	7	0.37	0.16	0.06	0.67	8.46	0.21
Yes	42	0.67	0.06	0.54	0.79	143.93	0.00
0 or 1 skills	13	0.34	0.10	0.15	0.53	15.98	0.19
2 or 3 skills	19	0.90	0.08	0.74	1.07	44.28	0.00
4 or more skills	17	0.54	0.08	0.38	0.71	50.60	0.00
Relations Domain							
No	2	1.53	0.33	0.89	2.17	0.26	0.61
Yes	47	0.60	0.06	0.48	0.71	143.79	0.00
0 or 1 skills	9	0.64	0.13	0.39	0.90	33.86	0.00
2 or 3 skills	29	0.55	0.07	0.41	0.69	71.21	0.00
4 or more skills	11	0.78	0.11	0.56	1.01	27.05	0.00
Operations Domain							
No	12	0.69	0.13	0.45	0.93	26.04	0.01
Yes	37	0.61	0.07	0.47	0.74	133.34	0.00
0 or 1 skills	23	0.74	0.09	0.56	0.91	69.27	0.00
2 or 3 skills	16	0.49	0.11	0.28	0.70	36.23	0.00
4 or more skills	10	0.60	0.12	0.37	0.82	44.13	0.00
Mathematical Language							
No	39	0.58	0.06	0.45	0.70	115.63	0.00
Yes	10	0.81	0.13	0.56	1.06	30.97	0.00

Note. Full sample (*N* = 51 effect sizes); outliers removed sample (*N* = 49 effect sizes). Within Group *Q* test represents heterogeneity within the variable noted (e.g., Number domain).

Appendix H
Numeracy Components

Table H1
Number Domain Skills

Study (year)	Cardinality	1-1 Counting[a]	Counting sequence[b]	Counting error[c]	Numeral ID	Subitizing
Bryant et al. (2011)			X		X	
Clarke, Doabler, Smolkowski, Baker, et al. (2016)	X	X	X		X	
Clarke, Doabler, Smolkowski, Kurtz Nelson, et al. (2016)	X	X	X		X	
Clarke et al. (2014)			X		X	
Clements (1984)	X	X	X			
Codding et al. (2011)	X	X	X		X	
Desoete & Praet (2013) [*compare*]						
Desoete & Praet (2013) [*count*]	X	X	X			
Doabler et al. (2016)	X	X	X		X	
Dyson et al. (2015) [*fact practice*]	X	X	X		X	X
Dyson et al. (2015) [*number list*]	X	X	X		X	X
Dyson et al. (2013)	X	X	X		X	X
Fuchs et al. (2005)			X		X	
Fuchs et al. (2001)	X	X	X		X	
Fuchs et al. (2002)					X	
Fuchs et al. (2013) [*speeded*]	X		X	X	X	
Fuchs et al. (2013) [*non-speeded*]	X		X	X	X	

Table H1 Continued…

Study (year)	Cardinality	1-1 Counting[a]	Counting sequence[b]	Counting error[c]	Numeral ID	Subitizing
Hansmann (2013) [*traditional QD*]						
Hansmann (2013) [*triangle QD*]						
Hassinger-Das et al. (2015) [*numeracy*]	X	X	X		X	
Hassinger-Das et al. (2015) [*story problems*]			X			
Jordan et al. (2012)	X	X	X		X	
Kaufmann et al. (2005)	X	X	X			X
Kyttälä et al. (2015)			X			
Mohanty & Mishra (1994)	X	X	X	X		
Obersteiner et al. (2013) [*exact*]						
Obersteiner et al. (2013) [*approximate*]						
Obersteiner et al. (2013) [*both*]						
Passolunghi & Costa (2016)		X	X			
Ramani & Siegler (2008)		X	X		X	
Salminen et al. (2015)		X	X			
Schacter et al. (2016)	X	X	X		X	X
Schopman & Van Luit (1996)	X	X	X			
Siegler & Ramani (2009) [*linear*]		X	X		X	
Sood & Jitendra (2011)	X	X				X
Toll & Van Luit (2012)		X	X		X	
Toll & Van Luit (2013)		X	X		X	

Table H1 Continued…

Study (year)	Cardinality	1-1 Counting[a]	Counting sequence[b]	Counting error[c]	Numeral ID	Subitizing
Van de Rijt & Van Luit (1998) [*guiding*]	X	X	X			
Van de Rijt & Van Luit (1998) [*structure*]	X	X	X			
Van Luit & Schopman (2000)		X	X		X	
Wilson et al. (2009)		X				

Note. Different treatment groups (by risk status, such as students with math difficulty versus typically achieving studies) were administered the same intervention; therefore, risk groups are not shown separately in this table; ID = identification; QD = quantity discrimination;
[a]1-1 Counting = Counting with 1-to-1 correspondence; [b]Counting sequence = oral counting, rote counting, verbal counting; [c]Counting error = Identifying errors in the counting sequence.

Table H2

Relations Domain Skills

Study	Magnitude comparison	Set comparison	NL estimation	NL sequence	Ordinal numbers	Matching quantity
Bryant et al. (2011)	X	X		X		
Clarke, Doabler, Smolkowski, Baker, et al. (2016)	X	X		X		X
Clarke, Doabler, Smolkowski, Kurtz Nelson, et al. (2016)	X	X		X		X
Clarke et al. (2014)	X			X		X
Clements (1984)		X				
Codding et al. (2011)	X	X		X		
Desoete & Praet (2013) [*compare*]	X	X				X
Desoete & Praet (2013) [*count*]						
Doabler et al. (2016)	X	X				X
Dyson et al. (2015) [*fact practice*]	X	X		X		
Dyson et al. (2015) [*number list*]	X	X		X		
Dyson et al. (2013)	X	X		X		X
Fuchs et al. (2005)	X			X		
Fuchs et al. (2001)	X	X		X		
Fuchs et al. (2002)	X			X		
Fuchs et al. (2013)	X			X		X
Hansmann (2013) [*traditional QD*]	X					
Hansmann (2013) [*triangle QD*]				X		

Table H2 Continued…

Study	Magnitude comparison	Set comparison	NL estimation	NL sequence	Ordinal numbers	Matching quantity
Hassinger-Das et al. (2015) [*numeracy*]	X	X		X		
Hassinger-Das et al. (2015) [*story problems*]						
Jordan et al. (2012)	X	X		X		
Kaufmann et al. (2005)		X				X
Kyttälä et al. (2015)	X	X	X	X		X
Mohanty & Mishra (1994)		X		X		
Obersteiner et al. (2013) [*exact*]				X		X
Obersteiner et al. (2013) [*approximate*]		X	X			X
Obersteiner et al. (2013) [*both*]		X	X	X		
Passolunghi & Costa (2016)		X		X		X
Ramani & Siegler (2008)				X		
Salminen et al. (2015)		X		X		X
Schacter et al. (2016)	X	X		X		X
Schopman & Van Luit (1996)		X				
Siegler & Ramani (2009) [*linear*]				X		
Sood & Jitendra (2011)		X		X		
Toll & Van Luit (2012)	X		X	X	X	
Toll & Van Luit (2013)	X		X	X	X	
Van de Rijt & Van Luit (1998) [*guiding*]	X	X			X	

Table H2 Continued…

Study	Magnitude comparison	Set comparison	NL estimation	NL sequence	Ordinal numbers	Matching quantity
Van de Rijt & Van Luit (1998) [*structure*]	X	X			X	
Van Luit & Schopman (2000)					X	
Wilson et al. (2009)	X	X		X		X

Note. Different treatment groups (by risk status, such as students with math difficulty versus typically achieving studies) were administered the same intervention; therefore, risk groups are not shown separately in this table. Magnitude comparison refers to the comparison of numerals (i.e., symbols), while set comparison refers to the comparison of sets of objects or quantities; NL = number line; QD = quantity discrimination.

Table H3
Operations Domain Skills

Study	Composing, decomposing, part-whole	Simple +/– (objects)[a]	Simple +/–[b]	Properties of +/–[c]	Place value, base-10
Bryant et al. (2011)	X		X		X
Clarke, Doabler, Smolkowski, Baker, et al. (2016)	X	X	X		X
Clarke, Doabler, Smolkowski, Kurtz Nelson, et al. (2016)	X	X	X		X
Clarke et al. (2014)		X	X	X	X
Clements (1984)		X			
Codding et al. (2011)	X	X	X		
Desoete & Praet (2013) [*compare*]					
Desoete & Praet (2013) [*count*]		X			
Doabler et al. (2016)	X	X	X		X
Dyson et al. (2015) [*fact practice*]	X		X		X
Dyson et al. (2015) [*number list*]	X		X		X
Dyson et al. (2013)	X	X	X		
Fuchs et al. (2001)	X	X	X		
Fuchs et al. (2002)		X	X		X
Fuchs et al. (2013) [*speeded*]	X	X	X	X	X
Fuchs et al. (2013) [*non-speeded*]	X	X	X	X	X
Fuchs et al. (2005)		X	X	X	X
Hansmann (2013) [*traditional QD*]					

Table H3 Continued…

Study	Composing, decomposing, part-whole	Simple +/– (objects)[a]	Simple +/–[b]	Properties of +/–[c]	Place value, base-10
Hansmann (2013) [*triangle QD*]					
Hassinger-Das et al. (2015) [*numeracy*]	X	X	X		X
Hassinger-Das et al. (2015) [*story problems*]					
Jordan et al. (2012)	X	X	X		X
Kaufmann et al. (2005)	X	X			
Kyttälä et al. (2015)					
Mohanty & Mishra (1994)		X			
Obersteiner et al. (2013) [*exact*]		X			
Obersteiner et al. (2013) [*approximate*]		X			
Obersteiner et al. (2013) [*both*]		X			
Passolunghi & Costa (2016)					
Ramani & Siegler (2008)					
Salminen et al. (2015)	X	X	X		
Schacter et al. (2016)					
Schopman & Van Luit (1996)					
Siegler & Ramani (2009) [*linear*]					
Sood & Jitendra (2011)	X	X			
Toll & Van Luit (2012)		X			
Toll & Van Luit (2013)			X		

Table H3 Continued…

Study	Composing, decomposing, part-whole	Simple +/– (objects)[a]	Simple +/–[b]	Properties of +/–[c]	Place value, base-10
Van de Rijt & Van Luit (1998) [*guiding*]					
Van de Rijt & Van Luit (1998) [*structure*]					
Van Luit & Schopman (2000)	X	X	X		
Wilson et al. (2009)			X		

Note. Different treatment groups (by risk status, such as students with math difficulty versus typically achieving studies) were administered the same intervention; therefore, risk groups are not shown separately in this table. [a]Addition and subtraction with objects; [b]Addition and subtraction without objects; [c]Properties of addition and subtraction.

Appendix I
Simple Comparisons of Effects by Numeracy Skill

Table I1
Simple Comparisons of Effects by Specific Number Domain Skills

Number Skill				95% Confidence Interval		Within Group		Between Groups	
	n	mean (*g*)	*SE*	Lower	Upper	*Q*	*p*(*Q*)	*Q*	*p*(*Q*)
Full Sample									
Cardinality	26	0.83	0.01	0.63	1.02	150.44	0.00	4.48	0.03
Correspondence	34	0.85	0.07	0.70	1.00	133.79	0.00	16.08	0.00
Counting Error	3	0.59	0.28	0.04	1.14	31.85	0.00	0.10	0.75
Count Sequence	38	0.74	0.08	0.59	0.90	176.71	0.00	2.85	0.09
Number ID	29	0.57	0.09	0.40	0.74	93.56	0.00	4.08	0.04
Subitizing	7	0.81	0.18	0.45	1.17	10.37	0.11	0.62	0.43
Outliers Removed									
Cardinality	24	0.72	0.09	0.56	0.89	90.28	0.00	2.49	0.12
Correspondence	32	0.79	0.06	0.67	0.91	78.51	0.00	19.90	0.00
Counting Error	2	0.25	0.23	-0.19	0.69	0.13	0.72	2.89	0.09
Count Sequence	36	0.68	0.07	0.55	0.81	117.00	0.00	2.51	0.11
Number ID	29	0.57	0.07	0.43	0.72	93.56	0.00	1.49	0.22
Subitizing	7	0.80	0.16	0.49	1.11	10.37	0.11	1.46	0.23

Note. Full sample (*N* = 51 effect sizes); outliers removed sample (*N* = 49 effect sizes). Within Group *Q* test represents heterogeneity within the variable noted (e.g., Cardinality); Across Groups *Q* test represents heterogeneity across variable coding of yes/no (i.e., the comparison between interventions that included Cardinality and interventions that did not include Cardinality). ID = identification.

Table I2

Simple Comparisons of Effects by Specific Relations Domain Skills

Relations Skill				95% Confidence Interval		Within Group		Across Group	
	n	mean (*g*)	*SE*	Lower	Upper	*Q*	*p*(*Q*)	*Q*	*p*(*Q*)
Full Sample									
Match Quantity	18	0.59	0.11	0.37	0.81	51.89	0.00	1.02	0.31
Missing Number	2	0.59	0.36	-0.10	1.29	2.44	0.12	0.06	0.81
NL Estimation	7	0.70	0.18	0.34	1.05	22.70	0.00	0.02	0.89
NL Sequence	36	0.59	0.08	0.43	0.74	130.82	0.00	4.44	0.04
Ordinal Numbers	7	0.97	0.17	0.63	1.31	12.38	0.05	3.35	0.07
Quantity Discrimination	33	0.60	0.08	0.44	0.76	111.16	0.00	3.11	0.08
Set Comparison	31	0.76	0.01	0.59	0.94	118.07	0.00	2.29	0.13
Outliers Removed									
Match Quantity	18	0.58	0.10	0.39	0.77	51.89	0.00	0.29	0.59
Missing Number	2	0.59	0.31	-0.02	1.21	2.44	0.12	0.01	0.92
NL Estimation	7	0.71	0.16	0.40	1.01	22.70	0.00	0.32	0.57
NL Sequence	35	0.56	0.07	0.43	0.69	101.47	0.00	3.29	0.07
Ordinal Numbers	7	0.97	0.15	0.68	1.26	12.38	0.05	6.59	0.01
Quantity Discrimination	33	0.59	0.07	0.45	0.73	111.16	0.00	0.68	0.41
Set Comparison	29	0.68	0.08	0.53	0.84	60.26	0.00	1.36	0.24

Note. Full sample ($N = 51$ effect sizes); outliers removed sample ($N = 49$ effect sizes). Within Group Q test represents heterogeneity within the variable noted (e.g., Match Quantity); Across Groups Q test represents heterogeneity across variable coding of yes/no (i.e., the comparison between interventions that included Match Quantity and interventions that did not include Match Quantity). NL = number line.

Table I3

Simple Comparisons of Effects by Specific Operations Domain Skills

Operations Skill				95% Confidence Interval					
	n	Mean (*g*)	*SE*	Lower	Upper	*Q*	*p*(*Q*)	*Q*	*p*(*Q*)
Full Sample									
Composing[a]	21	0.58	0.10	0.37	0.78	68.73	0.00	1.56	0.21
Equivalence	1	0.56	0.45	-0.32	1.44	0.00	1.00	0.07	0.79
Place Value	15	0.52	0.11	0.30	0.74	56.30	0.00	3.02	0.08
Properties of +/-	6	0.51	0.18	0.16	0.87	11.19	0.05	0.94	0.33
+/- with objects	32	0.65	0.09	0.48	0.82	162.34	0.00	0.30	0.58
+/- without objects	26	0.56	0.09	0.37	0.73	97.56	0.00	4.32	0.04
Outliers Removed									
Composing	21	0.58	0.09	0.41	0.76	68.73	0.00	0.44	0.51
Equivalence	1	0.56	0.38	-0.18	1.31	0.00	1.00	0.03	0.87
Place Value	15	0.52	0.09	0.33	0.70	56.30	0.00	2.09	0.15
Properties of +/-	6	0.50	0.15	0.21	0.79	11.18	0.00	0.79	0.37
+/- with objects	30	0.57	0.07	0.42	0.71	99.91	0.00	1.59	0.21
+/- without objects	26	0.55	0.08	0.40	0.71	97.56	0.00	1.93	0.17

Note. Full sample ($N = 51$ effect sizes); outliers removed sample ($N = 49$ effect sizes). Within Group *Q* test represents heterogeneity within the variable noted (e.g., Composing); Across Groups *Q* test represents heterogeneity across variable coding of yes/no (i.e., the comparison between interventions that included Composing and interventions that did not include Composing). +/- = addition and subtraction

[a] Composing = composition, decomposition, part-part-whole.

Appendix J

Table J1
Simple Comparisons of Effects by Study Information

Study Information				95% Confidence Interval		Within Group		Between Groups		
	n	mean (*g*)	*SE*	Lower	Upper	*Q*	*p*(*Q*)	*Q*	*df*	*p*(*Q*)
Full Sample										
Year of Publication								0.90	1	0.34
2012 or Later	29	0.62	0.09	0.45	0.80	116.10	0.00			
2011 or Earlier	22	0.76	0.11	0.55	0.97	104.44	0.00			
Location								6.11	1	0.01
United States	30	0.55	0.08	0.39	0.71	118.68	0.00			
Other	21	0.88	0.11	0.67	1.09	73.85	0.00			
Outliers Removed										
Year of Publication								0.01	1	0.91
2012 or Later	29	0.62	0.08	0.47	0.77	116.09	0.00			
2011 or Earlier	20	0.63	0.10	0.45	0.82	44.00	0.001			
Location										
United States	29	0.51	0.07	0.37	0.64	84.91	0.00	8.40	1	0.01
Other	20	0.84	0.09	0.66	1.02	47.55	0.00			

Note. Full sample ($N = 51$ effect sizes); outliers removed sample ($N = 49$ effect sizes). Within Group *Q* test represents heterogeneity within the variable noted (e.g., United States); Across Groups *Q* test represents heterogeneity across variable coding of (i.e., the comparison between Locations).

Appendix K

Table K1

Simple Comparisons of Effects by Methodological Characteristics

				95% Confidence Interval		Within Group		Between Groups		
Full Sample	*n*	mean (*g*)	*SE*	Lower	Upper	*Q*	*p*(*Q*)	*Q*	*df*	*p*(*Q*)
Treatment Effects										
Selected Treatment Groups Only	7	0.51	0.15	0.21	0.80	27.72	0.00	0.05	1	0.82
Non-selected Treatment Groups Only	7	0.46	0.15	0.16	0.75	12.94	0.04			
Pre- and Post-test	46	0.70	0.07	0.56	0.84	217.75	0.00	0.90	1	0.34
Post-test Only	5	0.49	0.21	0.08	0.90	2.67	0.61			
Assignment										
Random	44	0.64	0.07	0.50	0.78	206.26	0.00	2.08	3	0.56
Matching	5	0.93	0.22	0.50	1.36	4.30	0.37			
Not Random	1	0.98	0.50	0.01	1.96	0.00	1.00			
Not Reported	1	0.81	0.54	-0.24	1.87	0.00	1.00			
Nature of Treatment										
Supplants (full)	9	0.72	0.16	0.41	1.03	49.53	0.00	5.65	3	0.13
Supplements	18	0.65	0.11	0.43	0.87	43.45	0.00			
Combination	11	0.44	0.15	0.14	0.73	22.86	0.01			
Not Reported	13	0.93	0.15	0.64	1.22	89.85	0.00			

Table K1 continued…

Full Sample	*n*	mean (*g*)	*SE*	95% Confidence Interval Lower	Upper	Within Group *Q*	*p*(*Q*)	Between Group *Q*	*df*	*p*(*Q*)
Nature of Control										
BAU (time control)	19	0.67	0.11	0.45	0.89	66.79	0.00	7.05	4	0.13
BAU (no time control)	14	0.55	0.12	0.31	0.79	53.88	0.00			
Other math intervention	2	0.71	0.35	0.02	1.40	0.19	0.67			
Active control (non-math)	12	1.00	0.15	0.70	1.30	76.90	0.00			
Not reported	4	0.37	0.25	-0.12	0.86	7.91	0.05			
Study Quality										
0 or 1 indicators	8	0.92	0.19	0.54	1.30	77.50	0.00	2.13	2	0.34
2 or 3 indicators	21	0.69	0.11	0.48	0.90	86.98	0.00			
4 or 5 indicators	22	0.60	0.10	0.40	0.81	55.26	0.00			
General Confidence										
Very confident	31	0.69	0.09	0.52	0.86	121.09	0.00	0.06	1	0.80
Mostly confident	20	0.65	0.12	0.43	0.88	99.76	0.00			
Effect Size Confidence										
No estimation	48	0.67	0.07	0.53	0.81	218.94	0.00	0.09	1	0.77
Some estimation	3	0.76	0.28	0.21	1.30	0.96	0.62			

Table K1 continued…

Full Sample	*n*	mean (*g*)	*SE*	95% Confidence Interval Lower	Upper	Within Group *Q*	*p*(*Q*)	Between Group *Q*	*df*	*p*(*Q*)
Outliers Removed										
Treatment Effects										
Dependent	14	0.54	0.11	0.32	0.76	36.21	0.00	0.83	1	0.36
Independent	35	0.66	0.07	0.52	0.80	121.46	0.00			
Pre- and Post-test	44	0.64	0.06	0.52	0.77	156.95	0.00	0.63	1	0.43
Post-test Only	5	0.49	0.18	0.13	0.84	2.67	0.61			
Assignment										
Random	42	0.58	0.06	0.46	0.70	144.48	0.00	3.92	3	0.27
Matching	5	0.93	0.19	0.56	1.30	4.30	0.37			
Not Random	1	0.98	0.44	0.13	1.84	0.00	1.00			
Not Reported	1	0.81	0.48	-0.13	1.75	0.00	1.00			
Nature of Treatment										
Supplants (full)	9	0.71	0.13	0.45	0.97	49.53	0.00	3.09	3	0.38
Supplements	18	0.65	0.09	0.47	0.84	43.45	0.00			
Combination	11	0.43	0.13	0.17	0.68	22.86	0.01			
Not Reported	11	0.70	0.13	0.44	0.95	30.99	0.00			

Table K1 continued…

Full Sample	*n*	mean (*g*)	*SE*	95% Confidence Interval Lower	Upper	Within Group *Q*	*p*(*Q*)	Between Group *Q*	*df*	*p*(*Q*)
Nature of Control										
BAU (time control)	*19*	0.67	*0.10*	0.47	0.86	*66.79*	*0.00*	*3.20*	*4*	*0.52*
BAU (no time control)	*14*	0.56	*0.10*	0.35	0.76	*53.88*	*0.00*			
Other math intervention	2	0.70	0.31	0.10	1.31	0.19	0.67			
Active control (non-math)	10	0.76	0.14	0.49	1.03	20.10	0.02			
Not reported	4	0.35	0.22	-0.07	0.78	7.91	0.05			
Study Quality										
0 or 1 indicators	6	0.54	0.17	0.20	0.88	16.35	0.01	0.62	2	0.73
2 or 3 indicators	21	0.68	0.09	0.50	0.86	86.98	0.00			
4 or 5 indicators	22	0.60	0.09	0.43	0.78	55.26	0.00			
General Confidence										
Very confident	31	0.68	0.07	0.54	0.82	121.09	0.00	1.70	1	0.19
Mostly confident	18	0.52	0.10	0.32	0.72	37.66	0.00			
Effect Size Confidence										
No estimation	46	0.62	0.06	0.50	0.74	157.74	0.00	0.30	1	0.58
Some estimation	3	0.75	0.24	0.29	1.22	0.96	0.62			

Note. See Appendix A for definitions of each code and coding procedures. Full sample (*N* = 51 effect sizes); outliers removed sample (*N* = 49 effect sizes). Within Group *Q* test represents heterogeneity within the variable noted (e.g., random assignment); Across Groups *Q* test represents heterogeneity across variable coding of (i.e., the comparison between all assignment types). BAU = business as usual.

Appendix L

Table L1
Simple Comparisons of Effects by Participant Grade and Risk Status

				95% Confidence Interval		Within Group		Between Groups		
Participant Characteristic	*n*	Mean (*g*)	*SE*	Lower	Upper	*Q*	*p*(*Q*)	*Q*	*df*	*p*(*Q*)
Full Sample										
Grade								19.02	2	0.00
Preschool	8	1.10	0.16	0.78	1.41	63.84	0.00			
Kindergarten	30	0.75	0.08	0.59	0.90	72.77	0.00			
First Grade	13	0.32	0.11	0.11	0.53	16.00	0.19			
Risk Status								2.25	3	0.52
Low Risk	18	0.64	0.12	0.40	0.87	106.95	0.00			
Some Risk (SES)	6	0.77	0.19	0.40	1.15	3.98	0.55			
MD (Other)	16	0.78	0.12	0.55	1.01	78.00	0.00			
MD (≤ 25th pc.)	11	0.52	0.15	0.22	0.82	22.47	0.01			
Outliers Removed										
Grade								18.98	2	0.00
Preschool	6	0.84	0.14	0.58	1.12	11.67	0.04			
Kindergarten	30	0.74	0.07	0.62	0.87	72.77	0.00			
First Grade	13	0.32	0.08	0.15	0.48	16.00	0.19			
Risk Status								5.78	3	0.12
Low Risk	16	0.47	0.10	0.27	0.68	42.30	0.00			
Some Risk (SES)	6	0.77	0.16	0.45	1.09	3.98	0.55			
MD (Other)	16	0.77	0.10	0.58	0.96	78.00	0.00			
MD (≤ 25th pc.)	11	0.51	0.13	0.26	0.77	22.47	0.01			

Note. Full sample (N = 51 effect sizes); outliers removed sample (N = 49 effect sizes). Within Group Q test represents heterogeneity within the variable noted (e.g., Preschool); Across Groups Q test represents heterogeneity across variable coding of (i.e., the comparison between all grade levels). Low risk = sample of students who were not screened for MD including students identified as typically achieving and potential risk defined by as low socioeconomic status; MD (Other) = identified as having a mathematics difficulty by criteria other than performance at or below the 25th percentile; MD ($\leq$ 25th pc.) = identified as having mathematics difficulty by performance at or below the 25th percentile.

Appendix M

Table M1

Simple Comparisons of Effects by Intervention Characteristics

				95% Confidence Interval		Within Group		Between Groups		
Intervention Characteristic	*n*	Mean (*g*)	*SE*	Lower	Upper	*Q*	*p*(*Q*)	*Q*	*df*	*p*(*Q*)
Full Sample										
Duration										
Less than or equal to 8 weeks	28	0.74	0.1	0.55	0.93	126.00	0	0.90	1	0.34
More than 8 weeks	23	0.61	0.1	0.42	0.80	93.30	0			
Hours Instruction										
0–9 hours	21	0.66	0.11	0.45	0.87	54.79	0.00	2.157	2	0.34
10–19 hours	19	0.80	0.12	0.57	1.02	98.26	0.00			
20+ hours	11	0.54	0.13	0.28	0.80	53.21	0.00			
One-on-one	14	0.60	0.11	0.37	0.82	41.99	0.00	20.17	4	0.00
Small group	25	0.84	0.09	0.68	1.01	107.9	0.00			
Flexible	2	1.28	0.33	0.64	1.90	0.30	0.59			
Peer-assisted	8	0.25	0.16	-0.06	0.55	3.45	0.84			
Not reported	2	0.05	0.29	0.50	0.61	0.10	0.76			
Intervention Agent										
Researcher	20	0.57	0.11	0.36	0.78	52.62	0.00	2.751	5	0.74
Teacher	15	0.84	0.14	0.57	1.11	112.98	0.00			

Table M1 Continued…

Intervention Characteristic	*n*	Mean (*g*)	*SE*	95% Confidence Interval Lower	Upper	Within Group *Q*	*p*(*Q*)	Between Groups *Q*	*df*	*p*(*Q*)
Other school staff	4	0.76	0.23	0.32	1.20	10.66	0.01			
Mixed	1	0.57	0.45	-0.31	1.44	0.00	1.00			
Computer-assisted	10	0.70	0.16	0.38	1.02	26.59	0.002			
Not reported	1	0.52	0.53	-0.53	1.56	0.00	1.00			
Outliers Removed										
Duration										
Less than or equal to 8 weeks	26	0.64	0.09	0.01	0.48	0.81	66.13	0.08	1	0.77
More than 8 weeks	23	0.61	0.08	0.01	0.44	0.77	93.30			
Hours Instruction										
0–9 hours	21	0.65	0.09	0.47	0.83	54.79	0.00	0.84	2	0.66
10–19 hours	17	0.67	0.10	0.47	0.86	40.61	0.001			
20+ hours	11	0.54	0.11	0.32	0.76	53.32	0.00			
One-on-one	14	0.57	0.09	0.39	0.75	41.99	0.00	25.00	4	0.00
Small group	23	0.76	0.07	0.63	0.90	51.95	0.00			
Flexible	2	1.28	0.28	0.73	1.83	0.30	0.59			
Peer-assisted	8	0.25	0.13	-0.01	0.49	3.45	0.84			
Not reported	2	0.05	0.23	-0.40	0.50	0.10	0.76			

Table M1 Continued…

Intervention Characteristic	*n*	Mean (*g*)	*SE*	95% Confidence Interval: Lower	Upper	Within Group: *Q*	*p*(*Q*)	Between Groups: *Q*	*df*	*p*(*Q*)
Intervention Agent										
Researcher	20	0.56	0.09	0.38	0.74	52.62	0.00	1.38	5	0.93
Teacher	13	0.65	0.12	0.41	0.89	54.2	0.00			
Other school staff	4	0.76	0.19	0.39	1.13	10.66	0.01			
Mixed	1	0.57	0.37	-0.16	1.29	0.00	1.00			
Computer-assisted	10	0.69	0.14	0.41	0.97	26.59	0.002			
Not reported	1	0.52	0.47	-0.41	1.43	0.00	1.00			

Note. Full sample (*N* = 51 effect sizes); outliers removed sample (*N* = 49 effect sizes). Within Group *Q* test represents heterogeneity within the variable noted (e.g., Preschool); Across Groups *Q* test represents heterogeneity across variable coding of (i.e., the comparison between all grade levels). At risk (SES) = majority of the participants were identified as low socioeconomic status; MD (Other) = identified as having a mathematics difficulty by criteria other than performance at or below the 25th percentile; MD (≤ 25th pc.) = identified as having mathematics difficulty by performance at or below the 25th percentile.

Appendix N

Table N1

Simple Comparisons of Effects by Instructional Features

				95% Confidence Interval		Within Group		Between Groups		
Instructional Feature	*n*	Mean (*g*)	*SE*	Lower	Upper	*Q*	*p*(*Q*)	*Q*	*df*	*p*(*Q*)
Full Sample										
Explicit Instruction	48	0.64	0.07	0.51	0.78	186.72	0.00	5.88	1	0.02
No	3	1.48	0.34	0.82	2.15	26.93	0.00			
Modeling	20	0.61	0.10	0.41	0.81	70.17	0.00	0.69	1	0.41
No	31	0.73	0.09	0.55	0.90	144.70	0.00			
Guided Practice	22	0.64	0.10	0.45	0.84	66.50	0.00	0.26	1	0.61
No	29	0.71	0.10	0.53	0.90	153.33	0.00			
Independent Practice	30	0.67	0.09	0.50	0.84	137.09	0.00	0.013	1	0.91
No	21	0.69	0.11	0.47	0.90	78.29	0.00			
Scripted Lessons	36	0.59	0.08	0.43	0.74	107.94	0.00	4.79	1	0.03
No	15	0.92	0.13	0.66	1.17	104.32	0.00			
Corrective Feedback	28	0.60	0.09	0.42	0.78	96.24	0.00	1.67	1	0.20
No	23	0.78	0.10	0.57	0.98	118.24	0.00			
Reinforcement	21	0.51	0.11	0.30	0.73	43.29	0.01	3.74	1	0.05
No	30	0.78	0.09	0.61	0.95	171.99	0.00			
Behavior Plan	4	0.40	0.21	0.34	0.89	5.39	0.15	1.97	1	0.16
No	47	0.71	0.07	0.57	0.84	193.59	0.00			
Concrete Representations	34	0.61	0.08	0.45	0.77	129.94	0.00	2.06	1	0.15
No	17	0.82	0.12	0.58	1.05	81.46	0.00			

Table N1 Continued…

				95% Confidence Interval		Within Group		Between Groups		
Instructional Feature	*n*	Mean (*g*)	*SE*	Lower	Upper	*Q*	*p*(*Q*)	*Q*	*df*	*p*(*Q*)
Pictorial Representations	41	0.61	0.07	0.46	0.75	148.72	0.00	5.22	1	0.02
No	10	1.02	0.16	0.70	1.33	65.92	0.00			
CRA Framework	12	0.80	0.14	0.53	1.07	30.86	0.001	1.09	1	0.30
No	39	0.64	0.08	0.49	0.79	174.73	0.00			
Game-based	14	0.72	0.14	0.46	0.99	28.26	0.01	0.16	1	0.69
No	37	0.66	0.08	0.51	0.82	190.63	0.00			
Included Games	15	0.51	0.11	0.28	0.73	41.05	0.00	3.12	1	0.08
No	36	0.75	0.08	0.60	0.91	151.30	0.00			
Outliers Removed										
Explicit Instruction	47	0.62	0.06	0.50	0.74	157.86	0.00	0.28	1	0.60
No	2	0.80	0.34	0.13	1.47	1.53	0.22			
Modeling	20	0.6	0.09	0.43	0.78	70.17	0.00	0.11	1	0.74
No	29	0.64	0.08	0.49	0.80	85.75	0.00			
Guided Practice	22	0.63	0.09	0.47	0.80	66.50	0.00	0.02	1	0.88
No	27	0.62	0.08	0.45	0.78	93.19	0.00			
Independent Practice	29	0.63	0.07	0.48	0.77	107.90	0.00	0.00	1	0.96
No	20	0.62	0.10	0.43	0.81	47.84	0.00			
Scripted Lessons	36	0.58	0.07	0.45	0.72	107.94	0.00	1.45	1	0.23
No	13	0.75	0.12	0.52	0.97	47.15	0.00			
Corrective Feedback	28	0.60	0.08	0.44	0.75	96.24	0.00	0.35	1	0.55
No	21	0.67	0.09	0.49	0.84	59.78	0.00			

Table N1 Continued…

				95% Confidence Interval		Within Group		Between Groups		
Instructional Feature	*n*	Mean (*g*)	*SE*	Lower	Upper	*Q*	*p*(*Q*)	*Q*	*df*	*p*(*Q*)
Reinforcement	21	0.51	0.09	0.33	0.69	43.29	0.002	2.49	1	0.11
No	28	0.70	0.08	0.55	0.85	112.21	0.00			
Behavior Plan	4	0.40	0.17	0.07	0.73	5.36	0.15	2.07	1	0.15
No	45	0.65	0.06	0.54	0.77	18.59	0.00			
Concrete Representations	33	0.58	0.07	0.44	0.72	100.68	0.00	1.30	1	0.25
No	16	0.72	0.10	0.52	0.93	51.79	0.00			
Pictorial Representations	40	0.58	0.06	0.46	0.71	119.72	0.00	2.67	1	0.10
No	9	0.84	0.15	0.56	1.13	36.91	0.00			
CRA Framework	12	0.81	0.12	0.58	1.03	30.86	0.001	3.20	1	0.07
No	37	0.57	0.07	0.44	0.69	112.10	0.00			
Game-based	14	0.72	0.12	0.49	0.95	28.27	0.00	0.82	1	0.37
No	35	0.59	0.07	0.46	0.73	129.07	0.00			
Included Games	15	0.49	0.10	0.01	0.31	0.68	41.05	2.73	1	0.10
No	34	0.69	0.07	0.00	0.55	0.82	93.67			

Note. "No" indicates that the intervention/study did not include the instructional feature listed previous to the word, "No." Full sample (*N* = 51 effect sizes); outliers removed sample (*N* = 49 effect sizes). Within Group *Q* test represents heterogeneity within the variable noted (e.g., Explicit Instruction [yes]); Across Groups *Q* test represents heterogeneity across variable coding of (i.e., the comparison between interventions that included Explicit Instruction and interventions that did not include Explicit Instructions); CRA = concrete-representational-abstract framework.