The Report Committee for Zhina Shen certify that this is the approved version of the following report:

Effect of Manipulative-based Interventions on Improving Mathematics Computation Outcomes for Students with Mathematics Difficulties

Committee:

____
Supervisor: Bindu Viswanathan

____
Co-Supervisor: Layla Guyot

____
Sarah R. Powell

Effect of Manipulative-based Interventions on Improving Mathematics Computation Outcomes for Students with Mathematics Difficulties

By

Zhina Shen

Report

Presented to the Faculty of the Graduate School of The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics.

The University of Texas at Austin

May, 2023

Acknowledgements

I am very grateful to Dr.Bindu Viswanathan, Dr.Layla Guyot for their comments and feedback, and regular meeting with me. I appreciate their time commitment on this report.

I would also like to thank my doctor advisor Sarah Powell for all her professional suggestions and editing for this manuscript. She provided me all the content related help.

I would like to thank my husband Haiyue Xu for his support and help with children care.

I would like to thank all my other colleagues' s support.

Effect of Manipulative-based Interventions on Improving Mathematics Computation Outcomes for Students with Mathematics Difficulties

by

Zhina Shen

The University of Texas at Austin, 2023

SUPERVISORS: Bindu Viswanathan, Layla Guyot

In this report, we discussed the effect of manipulative-based interventions on improving mathematics computation outcomes for students with mathematics difficulties. We conducted systematic analysis on 23 articles including single case design studies, and group design studies by using random effect meta regression model. This study serves as the first study which combined single case design and group design studies in the data analysis in the field of manipulative-based interventions. We determined that the manipulative-based interventions are effective on improving whole number computation for students with mathematics difficulties.We also found that special education teacher or researcher implemented interventions had a larger effect size than that implemented by general education teachers. However, our conclusion should be considered as preliminary and more exploration on manipulative-based interventions are needed.

Table of Contents

Chapter 1:Introduction

Mathematics difficulty (MD) impacts students across mathematics content areas (Nelson & Powell, 2018). Students with MD have more difficulty than peers without MD in early numeracy skills such as counting, fact retrieval, number-set identification, mathematics fact fluency, place value, and understanding mathematics operation signs (Andersson, 2010; Cirino et al., 2015; Nelson & Powell, 2018). Similarly, for word-problem solving (WPS), students with MD, compared to their peers without MD, experience higher rates of difficulty in mathematics language and word-problem tasks (Arsenault & Powell, 2022; Kingsdorf & Krawec, 2014; Pongsakdi et al., 2020). Continuing to rational numbers, students with MD experience more difficulty with fraction and decimal representation, comparison, and computation (Nelson & Powell, 2018).

Computation represents one persistent area of need for students with MD (Andersson, 2010; Nelson & Powell, 2018). For this meta-analysis, I define computation as adding, subtracting, multiplying, or dividing single- or multi-digit whole numbers. Examples include adding single-digit numbers (i.e., $3 + 5 + 7 = \_$), subtracting with a multi-digit minuend and single-digit subtrahend (i.e., $24 - 7 = \_$), multiplying multi-digit numbers (i.e., $43 \times 22 = \_$), and dividing with a double-digit dividend and a single-digit divisor (i.e., $45 \div 5 = \_$). In the elementary grades, students with MD demonstrate lower accuracy, take longer, and use more inefficient strategies than their peers without MD when solving computation problems (Andersson, 2010; McCaskey et al., 2018; Nelson et al., 2018). Additionally, computation skills predict later algebraic readiness, and computational performance at first grade acts as a predictor for third grade performance on state testing (Kiss et al., 2019).

When developing computational skills, students must develop both procedural and conceptual knowledge. Procedural knowledge includes following steps to solve a problem (Rittle-Johnson & Alibali, 1999). For instance, when dividing a multi-digit number by a single-digit number, students with an understanding of procedural knowledge may solve the problem using the steps of long division. Conceptual knowledge is an understanding of the concepts in a domain and how those concepts interact (Rittle-Johnson & Alibali, 1999). For example, while adding two double-digit numbers, students may demonstrate conceptual understanding by solving the problem with a number line, using base-ten blocks, or explaining what it means to regroup (e.g., exchange 10 ones for 1 ten). To exhibit mastery with computation, students must develop procedural and conceptual understanding of single- and multi-digit addition, subtraction, multiplication, and division problems (Gilmore et al., 2016; Österman & Bråting, 2019). In this study, I focus on the conceptual piece of computation. Strong conceptual skills can both support students to build an understanding of how concepts interact and reduce student reliance on procedural skills (Gilmore et al., 2016). One method to support students to build a conceptual understanding for computation problems is through the use of manipulatives (Carbonneau et al., 2013; Hinton & Flores, 2019).

Definition of Manipulatives

Manipulatives are commonly used in three formats, including concrete manipulatives, virtual manipulatives, and concrete representational abstract (CRA) sequence. Concrete manipulatives refer to concrete materials or physical objects that students can manipulate to explore and develop an understanding of a mathematical concept, including number sense, computation, WPS, etc. Concrete manipulatives include but are not limited to the following: pattern blocks, base 10 blocks, tiles, cubes, fraction circles, square tiles, and plastic/paper money, etc., (Bouck & Flanagan, 2010; Bouck & Park, 2018;).Virtual manipulatives are modeled after concrete manipulatives, but they are delivered through a computerized tool or application. It has been defined as an interactive, web-based visual representation of a dynamic object that presents opportunities for constructing mathematical knowledge (Bouck et al., 2018; Moyer et al., 2002). More than the simple combination of concrete and virtual manipulatives, CRA is a an instructional technique that entails presenting students with representations of mathematics concepts that move from concrete to abstract in three stages (concrete-pictorial-formal symbols) (Lafay et al., 2019). It includes three phases: concrete, representational, and abstract. In concrete phase, students use concrete manipulatives to solve mathematic problems. In representational phase, students use drawings solve mathematics problems. In abstract phase, students use standard algorithm to solve word problems.

Definition of Mathematics Difficulties (MD)

In the literature, *mathematics disability* refers to a student with a school-identified learning disability with IEP (Individualized Education Program) goals related to mathematics. *Mathematics difficulty* is defined as low mathematics performance, and this is typically determined by a teacher or scores on a researcher-implemented screening test. In this study, authors of included studies identified mathematics difficulty in different ways. Primarily, authors used cut-scores (e.g., <25^{th} percentile) but some others used methods such as whether the student had a school-identified learning disability with Individualized Education Program (IEP) goals in mathematics. In the literature, this variability in identification of mathematics difficulty is common (Nelson & Powell, 2018). Based on above, we referred to students with mathematics disability or difficulty as students experiencing mathematics difficulties (MD).

Definition of Whole Number Computation (WNC)

We included studies that at least had one performance outcome as whole number computation (WNC), which is also the dependent variable in my study. We define whole number computation (WNC) as computations that involve whole numbers' addition, subtraction, multiplication, and division, which could be either single-digit computation or multiple-digit computation with or without regrouping. We excluded studies that only had other types of computation as their academic outcomes, e.g., fractional computation, algebraic computation, or decimal computation.

Chapter 2:Existing Systematic Review Studies

Systematic reviews on the use of manipulative-based interventions to improve mathematical outcomes for students with and without disabilities determined a similar conclusion: manipulatives are effective at improving mathematical outcomes (Bouck & Park, 2018; Bouck et al., 2018; Flores & Hinton, 2019; Lafay et al., 2019; Peltier et al., 2020). However, effects differed based on population and program characteristics. More research is needed on moderators which impact the effectiveness of manipulative-based interventions. Despite previous research investigated the impact of moderators on manipulative-based interventions, including students' age, disability, the type of manipulatives, implementers, etc., insufficient analyses and conflicting findings exist (Carbonneau, 2013; Domino, 2010; Holmes, 2013; Peltier et al., 2020; Shin et al., 2021). Based upon previous research and educational practice, we would like to explore more on the following moderators, methodology characteristics (study design, implementer, and study quality), content of measure (addition, subtraction, multiplication, and division), and instructional characteristics (grade level, proportion of males). We will extend previous moderation analyses from the following aspects.

First, none of the existing systematic reviews have done quantitative analysis by using comparable metric for single case design (SCD) and group design studies, which means, no quantitative analysis have been done to combine the Effect Sizes (ES) of SCD and group design studies. Peltier et al. (2020) and Shin (2021) conducted quantitative analysis of manipulative effects on mathematics performance, however, they both focused on SCD studies, and did not include group design studies. This study serves as the first meta-analysis to combine the ESs of SCD and group design studies' in the area of manipulative-based interventions research. By combining ESs of SCD and group design, we would also be interested in if the effects of manipulative-based interventions vary with study quality.

Second, we would compare the effectiveness of manipulative-based interventions among researchers, special education teachers, and general education teachers. However, previous meta-analyses only investigated whether the effect of manipulative-based interventions vary between researchers and teachers (Carbonneau et al., 2013; Peltier et al., 2020). Given that different populations of students special education teachers and general education teachers work with, it is necessary to investigate the efficacy of manipulative-based interventions separately between them.

Third, most of previous syntheses analyzed a wide variety of students' mathematical outcomes, e.g., word-problem solving, fractions, retention, transfer, etc. (Bouck et al., 2018; Bouck & Park, 2018; Carbonneau et al., 2013), however, we will only target on the outcome of whole number computation (WNC). Also, other researchers either focused on students in general education or students with different types of disabilities (e.g., autism spectrum disorder, emotional behavior disability, etc), however, we would focus on students with MD. This could provide teachers or researchers with more specific reference to mathematics teaching for students with MD.

Fourth, developmental theorists suggested that a child's developmental status would dictate how knowledge is learned (Bruner, 1964; Piaget, 1962). Prior meta-analyses yielded conflict finding regarding the impact of students' age level on the effectiveness of manipulative use. Carbonneau et al. (2013) investigated whether the use of manipulatives was differential across ages of participants (from kindergarten to college level). They concluded students in the age group of 7 to 11 years old benefited more from using manipulatives than students 12 years and older. Whereas in other meta-analyses, Shin et al. (2021) and Peltier et al. (2020) identified middle and high school students with disabilities showed higher average effect than elementary students when using manipulative-based interventions. Due to this conflicting conclusion, more studies on moderation analysis of grade level are needed.

Lastly, there has not been much research conducting moderation analysis of the impact of proportion of males on manipulative-based interventions. Due to the potential differences in developmental characteristics between males and females, it will be important to analyze if the proportion of different genders moderate the effectiveness of manipulative-based intervention. Holmes (2013) intended to conduct the moderation analysis of the percentage of males on manipulative use, however, they did not obtain quantitative result possibly due to small number of studies providing this information. More manipulative-based intervention studies published since then, and we would like to update this moderation analysis by including more data.

Focus of My Meta-analysis

Taken together, the evidence for the effects of using manipulatives to support student mathematics achievement exists for either group design studies or SCD. Despite an increasing number of group design studies and SCD on manipulative use, no published meta-analytic review that used a comparable metric to compare or combine the ESs of group design studies and SCD exists. However, some reviews reported the impact of some moderators, e.g., students' age, disability status, the type of manipulatives, or implementers on manipulative use (Carbonneau, 2013; Domino, 2010; Holmes, 2013; Peltier et al., 2020; Shin et al., 2021). More moderation analyses are still needed. I would like to extend previous moderation analyses to examine if effects could differ based on methodology characteristics (study design, implementer, and study quality), content of measure (addition, subtraction, multiplication, and division), and instructional characteristics (grade level, proportion of males).

Results from this meta-analysis will aid researchers and practitioners in intervention design, and estimate students' mathematics performance based on instructional features and students' demographic characteristics. The present meta-analysis sought to address the following questions by conducting a systematic review of interventions involving manipulative use for students with MD.

1. What is the effect of manipulative-based interventions on improving whole number Computation (WNC) outcomes for students with mathematics difficulties (MD)?
2. Do effects differ based on methodology characteristics (study design, implementer, and study quality)?
3. Do effects differ based on the content of the measure: addition, subtraction, multiplication, or division?
4. Do effects differ based on instructional characteristics (grade level, proportion of males)?

Chapter 3: Method

Search Procedures

I conducted a search of five electronic databases: Academic Search Complete, Educational Resources Information Clearinghouse (ERIC), PsycINFO, ProQuest, and Education Source. I limited the search from 1929 to January of 2022, which could allow me to get a complete list of existing and relevant articles.

I used the following search terms: *manipulat** OR *cubes* OR *counters* OR *blocks* OR *dice* OR *domino** OR *base ten* OR *hands on* OR *CRA* OR *concrete representation** OR *physical item** OR *physical material** AND *math** OR *computation* OR *addition* OR *subtract** OR *multiplication* OR *multiply* OR *division* OR *divide* OR *fractions* AND *math* disabil** OR *math* disorder** OR *learning disab** OR *learning disorder** OR *learning difficult** OR *intellectual disabil** OR *developmental disab** OR *students with disabilities* OR *special education* OR *at risk*

Figure 1 displays a PRISMA diagram describing the search process (Page et al., 2021). The initial search yielded 1,831 abstracts. After removing duplicates and marking ineligible abstracts flagged by *ZOTERA*, I identified 1,234 abstracts. I screened the 1,234 abstracts, and I identified 70 articles that met the initial inclusion criteria. Then, I reviewed the full text of these 70 articles using the specific inclusion criteria described below, and I identified 27 studies that met the inclusion criteria for this meta-analysis.

Beyond the electronic search, we conducted an ancestral search using the reference lists from relevant mathematics syntheses or meta-analyses (Bouck et al., 2022; Carbonneau et al., 2013; Gersten et al., 2009; Jitendra et al., 2016; Peltier et al., 2020). This ancestral search resulted in an additional 3 studies for review. Finally, we completed a hand search for the following journals: *Exceptional Children, Teaching Exceptional Children, Preventing School Failure, the Journal of Special education, Remedial and Special Education, Learning Disability*

Quarterly, Learning Disability Research and Practice, Journal of Learning Disabilities, Learning Disabilities. We selected these journals because they contained the most relevant empirical research in the field of special education. We identified 2 additional articles through the hand search. In total, we included a total of 32 articles in this meta-analysis. Ten of the studies were group design studies, and 22 of them were SCD studies.

Inclusion and Exclusion Criteria

To be included, studies needed to meet the following inclusion criteria: (a) published in English, (b) peer-reviewed article or dissertation, (c) experimental group design or single case design (SCD), (d) if single case design, at least 3 subjects for each effect size (ES), (e) included at least one outcome related to whole number computation (WNC), (f) included students had mathematics difficulties. We excluded studies that (1) included manipulatives as part of their intervention components, instead of the main focus, (2) had less than three subjects for each of the ESs, (3) missed critical data for ES calculation, (4) had graphs in which data points cannot be extracted based on our method (Rohatgi, 2014).

Methodological Evaluation

We evaluated each study's methodological quality using the Council of Exceptional Children's quality standards for classifying evidence-based practices in special education (Cook et al., 2015). Studies received a rating (i.e., 0 or 1) according to each of the group comparison indicators within the following areas: context and setting, participants, intervention agent, description of practice, implementation fidelity, internal validity, outcome measures, dependent variables, and data analysis. An overall quality rating (i.e., ranging from 0 to 1) was calculated by averaging a study's scores across all indicators.

Coding Procedure

We developed a codebook based on the research questions of this meta-analysis. The codebook included the following sections: (a) study design: group design or SCD; (b) implementation agent: researchers, general education teachers or special education teachers; (c) grade level: grade 1 to grade 8; (d) percentage of male participants; (e) math content: addition, subtraction, multiplication, and division; (f) types of manipulatives: concrete manipulatives, virtual manipulatives, or CRA; (g) calculated quality indicators.

The first author served as the initial coder for all codes. The first author trained the second coder (third author), a doctoral student in special education department, to use the coding rubric, and we resolved disagreements through discussion for training purposes. Before the second coder started to code articles, they had to reach 90% agreement on the practice articles. Following the training, the second coder independently coded 25% of the articles for a reliability check. We calculated interrater agreement as [agreements/(agreements + disagreements) × 100]. Across the total variable matrix, the mean interrater agreement was 0.97, with acceptable values for all codes as follows: 1.00 for types of manipulatives, 1.00 for math content, 0.95 for percentage of male participants, 0.96 for study design, 1.00 for grade level, 0.98 for implementation agent, and 0.95 for quality indicators. The coders resolved any disagreements between raters by discussion.

Chapter 4: Analytic Strategies

Group Design Studies

The effect size index used for all group design studies' outcome was Cohen's *d*. Due to very similar study design, measure and sample, we tested the mean effect size estimate across all group design studies by using fixed effects model initially (Hedges, 1994). However, the Q-statistic was statistically significant ($p < 0.05$), which means the homogeneity of the sample was rejected. This further means the source of this variability in effect size estimates could be due to study sample characteristics rather than only sampling error. Due to the above rationale, we determined to use random effects model (Lipsey & Wilson, 2001). For studies reporting multiple effect sizes from the same sample, we accounted for the statistical dependencies by using the random effects robust variance estimation (RVE) technique developed by Hedges et al. (2010). Additionally, in order to correct the small standard error in random effects model, we adopted the random effects model with Knapp-Hartung correction technique (Knapp, 2003).

Single Case Design (SCD) Studies

Most of the effect size in SCD studies were expressed as nonoverlap (PND) or Tau-*U,* while a major limitation to these two ESs are the inability to account for the magnitude. Hedges et al. (2012, 2013; Shadish et al., 2014) introduced BC-SMD (between case standard mean difference) that is analogous to Cohen's *d* but it takes the magnitude of change into account, and is comparable with group design ESs. Similarly, in our study, to make the group design and SCD studies' ESs comparable, first, we digitized the data from graphs in SCD studies using procedures reported elsewhere that have excellent reliability and validity (https://automeris.io/WebPlotDigitizer/; Rohatgi, 2014). Then we used a web-based calculator (https://jepusto.shinyapps.io/scdhlm/; Pustejovsky, 2016) to calculate the BC-SMD d-statistic. In this study, we call ESs of both Cohen's *d* or BC-SMD *d*-statistics for the purpose of convenience and consistency.

Rosner's Test

An outlier is an observation that appears to deviate markedly from other observations in the sample. Based on examination of calculated *d*-statistics, there might be potential outliers in both group design and SCD studies. To detect outliers, we performed *Rosner*'s test in R for group design, and SCD studies separately (Rosner, 1975). Then, we excluded outliers from data analysis process, which included main effect analysis, and moderation analysis, to make our analysis result more reliable. After the Rosner's test, we excluded 2 studies, and the number of our final studies included in analysis is 21.

Moderation Analysis

The significant Q-test ($p < 0.05$) for heterogeneity indicates that effect sizes differ from each other by significantly more than would be expected due to chance alone. To account for the significant variation between studies, we examined the following moderators using random effects meta regression model with robust variance estimation (RVE) by combining group design and SCD studies: methodology characteristics (i.e. study design, implementer, and study quality), content of measure (i.e., addition, subtraction, multiplication, and division), instructional characteristics (i.e., grade level, and proportion of male) as shown in Table 3. The categorical variables included study design, implementer, and content of measure, whereas the numerical variables included study quality, proportion of male, and grade level.

We used a separate univariate random effects meta regression model for each moderator: study type (SCD or group), type of educator (researcher, special education teacher, general education teacher), measure (addition, subtraction, multiplication, division), study quality, proportion of male, and grade level. All numerical predictors (study quality, proportion male, grade level) were mean-centered. No alpha adjustment was used.

Chapter 5: Preliminary Results

Preliminary Analysis

A total of 33 studies met the inclusion criteria, including 11 group design studies and 22 single case design studies (see Table 1). As noted in Table 2, we excluded 10 studies from data analysis. Within the 10 excluded studies, we could not calculate effect sizes for 8 of the studies (see Table 2 for reasons). We also excluded 2 studies which were determined as outliers by Rosner's test.

Final Sample Used in the Analysis

We included 23 studies for data analysis, which comprised 6 group design studies, and 17 single case design studies. With these 23 studies, we calculated 32 effect sizes including 23 effect sizes (*d* statistics) for single case design studies, and 9 effect sizes (Cohen's *d*) for group design studies. These studies spanned 28 years from 1993 to 2021.

Effectiveness of Manipulative-based Interventions on Whole Number Computation (WNC)

In Table 1, we presented the effect size (Cohen's *d*) for group design studies, and the effect size (between-case standardized mean difference, or *d* statistic) for single case design studies. The effect size for group design studies ranged from -0.4 to 3.7 and the effect size for single case design studies ranged from -0.03 to 24.2. The average effect size for group design studies was 1.356 and significant ($p < 0.05$). After we performed Rosner's test for single case design studies' and group design studies' effect sizes, two outliers were detected from single case design studies and no outliers were detected from group design studies. The average effect size for single case design after removing two outliers was 2.518 ($n = 23$, $p < 0.05$). The average effect size for all studies was 2.014, and significant ($p < 0.05$).

Testing for Moderators on the Overall Effect Sizes

We summarized the findings from the analysis of the moderator variables in Table 3. We examined the moderators of methodology characteristics (i.e., study design, implementer, and study quality), content of measure and instructional characteristics (i.e., proportion of male, grade level). For each of the moderators, we presented effect size for each of the categories, standard error (SE), and 95% confidence interval (CI). We further presented the *p*-value after conducted the comparison analysis between categories within each of the moderators.

Methodology Characteristics

We examined three key methodology characteristics including study design, implementer, and study quality. Both the group design studies and single case design studies yielded large effect sizes, 2.518 and 1.356 respectively. The contrast between the mean effect size estimate (ES = 2.518) for single case design studies and that (ES = 1.356) for group design studies was not statistically significant, suggesting that the study design did not moderate the effectiveness of manipulative-based interventions. Regarding implementer, there was a significant difference ($p < 0.05$) when studies implemented by general education teachers, special education teachers and researchers. The mean effect size estimate (ES = 1.658) was largest for researcher-implemented studies, which was also similar to the mean effect size estimate (ES = 1.314) of special education teacher-implemented studies. The mean effect size estimate (ES = 0.339) of general education teacher-implemented studies was significantly smaller than those of researcher implemented studies and special education implemented studies. But this result also needs further exploration given that students' age are not very clear in each of those studies.The mean effect size estimate of study quality was large (ES = 4.31) and significant ($p < 0.0001$), which indicated the quality of studies moderated the effectiveness of manipulative-based interventions significantly.

Content of Measure

There was significant difference ($p < 0.05$) between content of measures: addition, subtraction, multiplication, and division. Among those four operations, addition (ES = 0.463) and division (ES = 0.899) had smaller effect sizes, whereas multiplication (ES = 1.38) and subtraction (ES = 1.782) had larger effect sizes. The mean effect size estimate of subtraction, multiplication and division were significantly larger than that of addition, suggesting that the content of measure moderated effectiveness of manipulative-based interventions.

Instructional Characteristics

We examined two instructional characteristics: grade level and proportion of male. The results showed grade level had large impact (ES = 0.732) on WNC, and this impact was significant ($p < 0.0001$). Also, as shown in Figure 1., as grade level increased, the estimated effect size of WNC also increased, and this impact was significant ($p < 0.05$). The proportion of male had large (ES = 0.773), but not significant impact ($p > 0.05$) on the WNC. As the proportion of male increased, the effectiveness of manipulative-based intervention on the WNC also improved.

Examining Publication Bias

We computed *Rosenthal's fail-safe N* (Begg & Mazumdar, 1994; Rosenthal, 1979) to assess the possibility of publication bias in the current meta-analysis. *Rosenthal's fail-safe N* provides an estimated number of studies that would be needed to make the effect insignificant. To assess the potential of publication bias, we calculated an estimate of the desired *fail-safe N*: 5 × (studies included in review) + 10 (Becker, 2005; Rosnow & Rosenthal, 1989). In the current meta-analysis, we computed desired *fail-safe N* as: (5 × 23) + 10 = 125. The obtained *fail-safe N*

(n = 664) based on the *fail-safe N* test in R was larger than the desired *fail-safe N*, thus we conclude publication bias is not a concern in our meta-analysis.

Chapter 6: Discussion

The primary goal of this meta-analysis was to examine the effectiveness of the manipulative-based interventions on WNC for students with MD. We also explored the moderation effects of methodology characteristics (i.e., study design, implementer, study quality), content of measure (i.e., addition, subtraction, multiplication, division), and instructional characteristics (i.e., grade level, proportion of male). We identified a limited number of studies in the area of manipulative-based interventions employing the same statistical metric (*d*-statistic) to analyze the combined ES of SCD and group design studies (e.g., Carbonneau et al., 2012). The overall ES ($d = 2.014$) suggested manipulatives were effective at improving the WNC performance of students with MD.

Methodology Characteristics

Study Design

Our preliminary results indicated that study design did not moderate the effectiveness of the manipulative-based interventions, meaning the effectiveness of manipulative-based interventions did not differ significantly between group design studies and SCD studies. This result was contrary to findings from Carbonneau et al. (2013) in which they determined SCD studies had a significantly higher mean effect size ($d = 1.22$) than group design studies ($d = 0.22$). This moderator difference could be attributed to several factors. First, the mathematics outcomes in Carbonneau et al. (2013) differed from the outcomes we measured in the present study. They coded their mathematics outcomes as retention (i.e., an outcome that required students to solve mathematics facts), problem solving, transfer, and justification. However, we focused specifically on mathematics whole number computation (WNC), which may overlap to some extent only with their retention outcome. Second, Carbonneau and colleagues (2013) included participants from kindergarten to college level with or without disabilities; we included only students with MD in order to provide recommendations for teaching this population of students. Even though we did not find significant difference of the mean ES between SCD and group design studies, we still determined SCD had larger mean ES (d = 2.5) than that in group design studies (d = 1.4), which was similar as Carbonneau et al. (2013). More review studies which incorporate both SCD and group design studies are needed to further verify study design differences.

We identified broad ES magnitude for SCD studies, which was similar as Peltier et al. (2020). Peltier and colleagues evaluated the SCD of 53 studies that implemented a manipulative-based intervention on the mathematical outcomes of students with disabilities. They found the ES (between case standardized mean difference or BC-SMD or d-statistic) ranged from 0.03 to 18.58 for individual SCD studies, whereas our ES (BC-SMD or d-statistic) for SCD studies ranged from -0.03 to 24.23. Even though Peltier et al. (2020) included broader mathematical outcomes (e.g., algebra, early numeracy, problem-solving), they also examined WNC as one of the mathematical outcomes. Therefore, there was still practical meaning to compare the results of our meta-analysis with results in their meta-analysis (Peltier et al., 2020).

As determined in Peltier et al. (2020) and our current meta-analysis, the range of ES for SCD studies were quite large, which could be caused by the small sample size in SCD studies. As shown in the following formula of d-statistic (Hedges et al., 2012), the numerator is the difference between the mean of treatment and control groups. For SCD studies, baseline data were used as the control group with intervention used as treatment group. The very large difference μ_d between baseline data and intervention data could cause the inflation of the SCD effect size (Busse et al., 1995; Zelinsky & Shadish, 2018).

$$\delta = \frac{\mu_d}{\sigma_d}$$

(where μ_d denotes the mean difference, and σ_d is the standard deviation of mean difference). The broad ES range in Peltier et al. (2020) and our study indicated the effect of manipulative-based interventions could vary substantially when implemented with small group or individual student through the lens of SCD. However, the overall mean ES of SCD was large ($d = 2.518$), and significant ($p < 0.05$), which suggested intervention is effective.

Implementer

In our preliminary results, we identified large and similar effect of researcher-implemented studies (ES = 1.658) and special education teacher-implemented studies (ES = 1.314). The mean effect size (ES = 0.339) of general education teacher-implemented studies was small, and significantly smaller than those of researcher-implemented studies, and special education teacher-implemented studies. It is reasonable to expect that researchers or special education teachers tend to implement manipulative-based interventions as planned more than general education teachers.

Researchers know their intervention components well, which made implementation fidelity of their studies higher than those done by other implementors (Bos et al., 2022). O'Donnell (2008) also determined that academic outcomes were significantly higher when the program was implemented with greater fidelity, which could explain the large effect size of researcher-implemented studies. Berkeley et al. (2010) conducted a meta-analysis analyzing the effect of reading comprehension instruction on improving reading ability for students from K to 12 grade level with learning disabilities.They determined higher outcomes for interventions implemented by researchers than teachers. The result of Dennis et al. (2016) showed that the interventions were much more effective when they were provided by the researchers or researcher-trained graduate assistants, however, interventions provided by teachers or paraprofessionals produced much weaker intervention effects.

Our moderation analysis showed that special education teacher-implemented studies had a similar effect as those implemented by researchers. However, special education teacher-implemented studies had larger effect than general education teacher-implemented studies. Potential explanations for this finding could include different types of teacher training programs in general education and special education or the differences in the populations of students these teachers teach. This might also be explained by the fact that the different studies were based on students of different age groups, and also used different ways of assessing MD. Special education teachers likely received professional development primarily designed for students with learning differences, which may have made these teachers more fluent with teaching tools, including manipulatives. Whereas, general education teachers received teacher training mainly focused on general education students, which may have made them less familiar with teaching method for special education students.

Our moderation result was consistent with Berkeley et al. (2010) and Dennis et al. (2010) in that researcher-implemented studies had higher intervention outcomes than those implemented by teachers. Because Berkeley et al. (2010) and Dennis et al. (2010) did not separately examine the category of teachers by special and general education teachers, we cannot compare our results of those from previous studies. But prior research results were consistent with ours on that researcher-implemented interventions tend to have a larger effect. More studies are needed to compare the effect between special education teacher-implemented studies and general education teacher-implemented studies. This could provide future researchers with information on how to train implementors, and the way to complete fidelity checks to ensure high implementation fidelity.

Study Quality

Our result showed that quality of studies moderated the effectiveness of manipulative-based interventions significantly. This result is inconsistent with Peltier et al. (2020), in which they learned study quality did not moderate intervention effects. One reason for this inconsistency could be that they only included SCD studies, whereas we included both group design studies and SCD studies. Peltier et al. (2020) evaluated their studies' quality based on What Works Clearinghouse standards (WWC; 2017), but we measured our studies' quality by using Cook et al. (2015; CEC, 2014). This application of different quality indicators could potentially interpret the effect of study quality differently.

Content of Measure

Our preliminary results showed that manipulative-based interventions had significantly larger effects on multiplication, division, and subtraction than addition. To date, we identified no meta-analyses in which the authors conducted a comparison analysis of the effect of manipulative-based interventions among these four operations. Thus, we do not have previous research as a reference. However, we could interpret our result based on the strategies needed for the four operations. For example, the effect of manipulative-based interventions on multiplication was larger than that for division. This conclusion is supported by the intuitive model (which was defined as an internal mental structure corresponding to a class of calculation strategies) studied by Mulligan and Mitchelmore (1997). They suggested students used three main intuitive models (i.e., direct counting, repeated addition, and multiplicative operations) in solving multiplication problems, whereas students use four models for division (i.e., direct counting, repeated subtraction, repeated addition, and multiplicative operations). Thus, it is reasonable to assume manipulative-based interventions only need to target three models for multiplication problems. In comparison, the manipulative-based interventions would need to focus on four models in order to help students improve their performance on division problems. This could potentially explain why the manipulative-based intervention had a stronger effect on multiplication than on division.

As described, the effect of manipulative-based intervention on addition was smallest, which could be due to the higher academic score of addition in pretest (Miller & Kaffar, 2011). Students may have already had a stronger understanding of solving addition problems than other operations before intervention. It is likely that students did not need as much practice on addition because they were introduced to learn addition earlier than any other operations starting from kindergarten (Common Core State Standards [CCSS], 2010). Thus, they had been working on addition the longest in schools, and already had stronger addition knowledge.

Instructional Characteristics

We determined grade level had a significant and large impact on WNC. As grade level increased, the effect size of WNC also increased. The trend identified in our meta-analysis was consistent with the review studies conducted by Moyer-Packenham and Westernskow (2013) and Shin et al. (2021). Shin et al. (2021) recognized the average immediate effect of virtual manipulatives was significantly greater for high school students than for elementary school students. Even though we only included participants in Grades 1 through 8 instead of participants in Kindergarten through Grade 12, the trend in Shin et al. (2021) and our study was similar. For the purpose of comparison in different age groups, more moderation studies are needed to analyze the effect of manipulative-based interventions in elementary, middle and high school level. We only analyzed six group design studies for the moderators of grade level and proportion of male. More group design studies are needed in order to further study the grade level and gender effect on the WNC performance.

Chapter 7: Limitations and Future Research

Although our findings suggest manipulatives are effective at improving WNC for students with MD, there are limitations that need to be considered when interpreting our results. First, we were not able to include all data from all included studies even when the author team provided all their raw data. The reason for this was provided in Table 2. For example, we excluded Hinton and Flores (2019) from ES calculation process. The reason for this was that we adopted Pustejovsky (2016) method (https://jepusto.shinyapps.io/scdhlm/) to calculate our SCD mean effect size for each of the studies. This method required there must be at least two repeated subjects for each ES, thus Hinton and Flores (2019) was excluded from the ES calculation process since there were only two subjects in their study. Future research may need to explore other ES calculation methods in order to incorporate more studies.

Second, for the moderators of proportion of male and grade level, we only conducted this analysis with group design studies. Both of these two moderators were considered as continuous variables, which could be either 0% or 100% in SCD studies. Thus, in SCD studies, it would lead to misinterpretation of the continuous moderators. Other techniques may be needed to include both SCD and group design studies for continuous moderators.

Third, there are also some statistical concern in this research. ES from SCD and group design studies are treated equally in this study. However, the ES from SCD studies does have inflation considering the small number of participants. However, the adjustment of the ES inflation is not adjusted in this study. Future studies may need to address this more by using other statistical technique. It might be more appropriate to weigh the group design studies' ESs higher than those from SCD studies. All models were univariate models, so we only assessed the effect of each moderator by itself. We did not do alpha adjustment. Thus, some of the significant results could be the product of a type I error. We also need to extend this research by applying models with multiple predictors taken together.

Chapter 8: Conclusions and Implications

Results from our current investigation hold important implications for educational practice. However, our conclusions should be considered as preliminary results based on all above limitations, and it needs to be explored more in the future. First, our results showed the effect of manipulative-based interventions were large and significant, which corroborated with other empirical studies, suggesting the manipulatives would improve the WNC outcomes for students with MD.

Second, there was no significant difference regarding the implementation of manipulative-based interventions in SCD studies and group design studies, suggesting manipulative-based interventions could be either implemented with individual students or in whole-class settings. A noteworthy point was that this meta-analysis study served as one of the first meta-analyses in the area of manipulative-based intervention research that combined the different ES metrics in SCD and group design studies by adopting the *d*-statistic calculation method. The combination of metrics could provide teachers more applicable reference in both individual level and group level.

Third, our results showed that researchers or special education teachers-implemented studies produced a significantly higher effect than general education teachers. This conclusion suggested there might be a gap between research and educational practice on manipulative-based interventions. More professional development on manipulative-based interventions may be needed for general education teachers in the future. Third, the manipulative-based interventions may be more effective on multiplication, division, and subtraction than addition, which could be explained by the intuitive model we referred to in the discussion section (Mulligan and Mitchelmore, 1997). However, this result should not be interpreted as not using manipulatives

with addition computation. For the addition computation, we could see that the ES for addition is still medium to large (Cohen et al., 2013). Therefore, teachers could implement manipulative-based interventions for all operations. With our analysis result in mind, they may expect a better intervention effect on multiplication, division, and subtraction. Finally, as grade level increased, the effect of manipulative-based interventions also increased. Our participants consisted of students from Grade 1 through Grade 8, which could provide a reference for elementary and secondary school teachers. Teachers should also be careful of the type of manipulatives used by different grade level given that virtual manipulatives can provide less stigmatization than concrete manipulatives, and can also be more socially desirable (Bouck et al., 2018).
As students enter higher grade levels, virtual manipulatives might be more appropriate.

Chapter 9: Future Work

In the future, we would like to extend this research more to make it more conclusive work. We would like to include descriptive analysis for all the moderators, analyzing ES based on group design and SCD separately. We would also like to extend to models with multiple predictors to investigate the effect of each moderator while controlling for other moderators. Also, by including more moderators in one model, we can also analyze interaction effect. We would also like to do alpha adjustment to ensure that all significant findings are interpretable.

Chapter 10: Appendix

Table 1. Effect Sizes for Group Design Studies and Single Case Design Studies

Studies	Year	Study Design	BC-SMD (*d* statistic)	Cohen's *d*
Milton et al.	2019	SCD	1.633	-
Milton et al.	2019	SCD	0.633	-
Milton et al.	2019	SCD	1.226	-
Hinton and Flores	2019	SCD	NC	-
Hinton and Flores	2019	SCD	NC	-
Flores and Milton	2020	SCD	0.400	-
Flores and Milton	2020	SCD	3.540	-
Flores, Hinton and Strozier	2014	SCD	0.761	-
Flores, Hinton and Strozier	2014	SCD	2.056	-
Flores, Hinton and Strozier	2014	SCD	1.672	-
Flores, Hinton and Schweck	2014	SCD	0.200	-
Flores	2009	SCD	1.040	-
Mancl, Miller and Kennedy	2012	SCD	6.262	-
Flores	2010	SCD	3.359	-
Flores	2011	SCD	1.997	-
Miller, Peterson and Mercer	1993	SCD	0.501	-
Miller, Peterson and Mercer	1994	SCD	-0.030	-
Derderian	2013	SCD	1.027	-
Carmack	2011	SCD	NC	-
Ferreira	2009	SCD	4.353	-
Harris, Miller and Mercer	1995	SCD	NC	-
Sealander et al.	2012	SCD	0.457	-
Wisniewski et al.	2002	SCD	NC	-
Bouck et al.	2020	SCD	4.641	-
Park et al.	2021	SCD	7.818	-
Park et al.	2020	SCD	4.805	-
Bouck, Shurr and Park	2020	SCD	6.540	-
Bouck, Park and Stenzel	2020	SCD	5.171	-
Bouck and Park	2020	SCD	12.765	-
Bouck, Long and Jakubow	2021	SCD	24.230	-
Flores, Moore and Meyer	2019	Group	-	3.297
Gibbs, Hinton and Flores	2018	Group	-	0.917
Flores and Franklin	2014	Group	-	1.231
Bennett and Rule	2005	Group	-	NC

Flores, Kaffar and Hinton	2019	Group	-	3.672
Tournaki, Young She and Kerekes	2008	Group	-	2.208
Miller and Kaffar	2011	Group	-	1.031
Miller and Kaffar	2011	Group	-	0.511
Miller and Kaffar	2011	Group	-	-0.392
Miller, Mercer and Dillon	1992	Group	-	NC
Miller et al.	1998	Group	-	NC
Smith and Montani	2008	Group	-	0.311
Smith and Montani	2008	Group	-	1.106

Note: NC stands for not calculable

Table 2. Studies Excluded from Data Analysis.

Studies	Reasons for incalculable effect sizes
Hinton and Flores	Only two subjects for each measured outcome
Bennett and Rule	Missing table for variance data
Tournaki, Young She and Kerekes	Correlational study
Carmack	Descriptive data
Harris, Miller and Mercer	Too many missing data points in single case design graph
Miller, Mercer and Dillon	No variance data provided
Wisniewski et al.	Descriptive data
Miller et al.	Mean of pre-test and post-test missing, no enough data
Bouck and Park	Outlier by Rosner’s test
Bouck, Long and Jakubow	Outlier by Rosner’s test

Table 3. Testing for Moderators of Effects Sizes Based on Random Effects Models.

Moderators	average effect size	SE	95%CI	p
All studies	1.227***	0.066	[1.099, 1.356]	
Methodology characteristics				
Study design (n = 23)				0.065
Group design(n=6)	1.356	0.608	[0.119, 2.593]	
SCD design(n=17)	2.518	0.739	[-0.343, 2.668]	
Implementer (n = 27)				<0.0001
Researcher(n=11)	1.658	0.192	[0.943, 1.695]	
Special education teacher(n=10)	1.314	0.175	[0.632, 1.319]	
General education teacher(n=6)	0.339	0.149	[0.046, 0.631]	
Study quality (n = 23)	4.31	0.715	[2.91, 5.71]	<0.0001
Content of measure (n = 40)				<0.0001
Addition(n=6)	0.463	0.173	[0.123, 0.801]	
Subtraction(n=11)	1.782	0.213	[0.903, 1.737]	
Multiplication(n=15)	1.38	0.21	[0.505, 1.330]	
Division(n=8)	0.899	0.213	[0.018, 0.854]	
Instructional characteristics				
Proportion of male (n = 6)	0.773	0.669	[-0.538, 2.084]	0.248
Grade level (n = 18)	0.732	0.052	[0.630, 0.835]	<0.0001

Figure 1
PRISMA Diagram (Page et al., 2021)

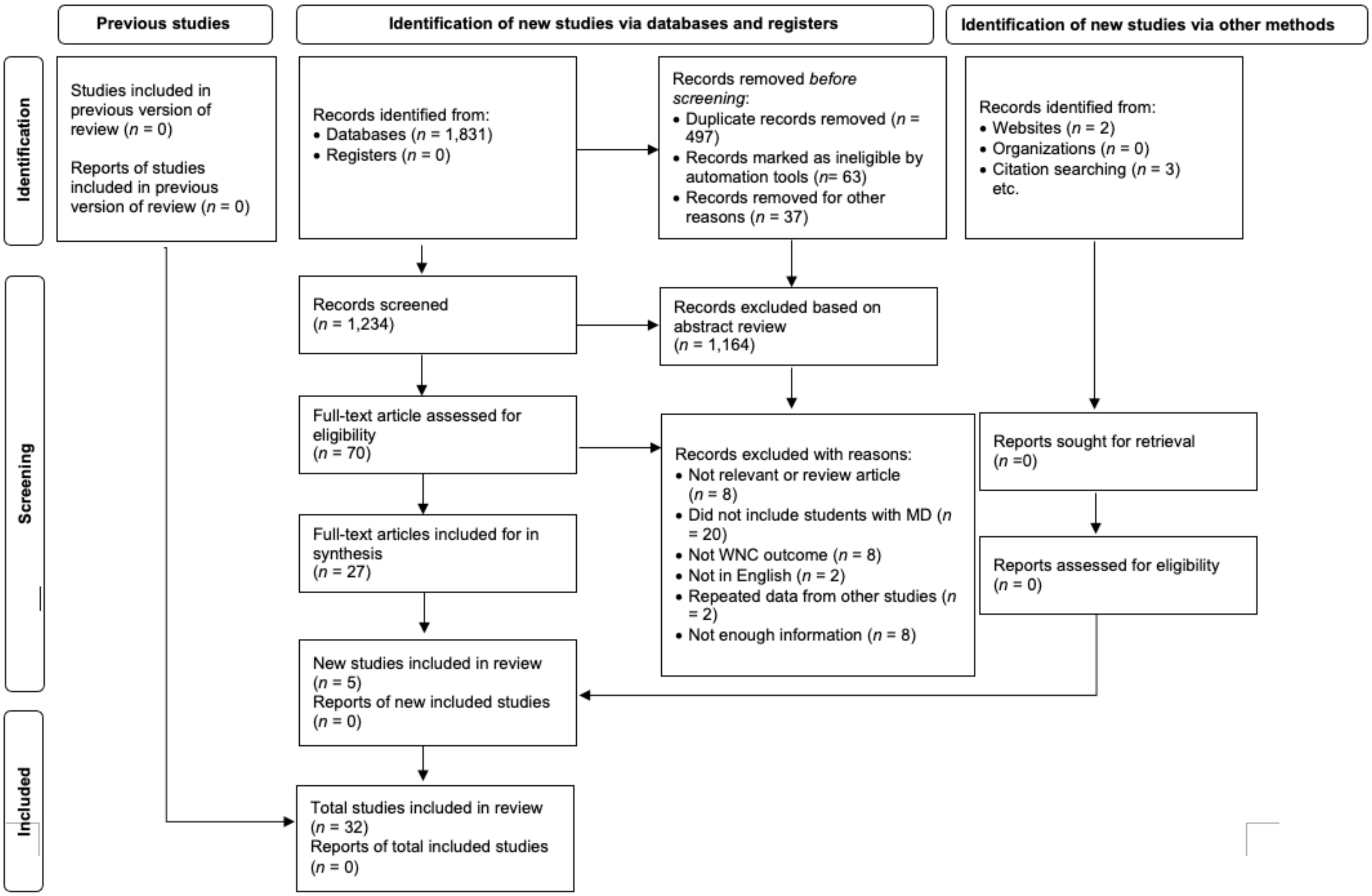
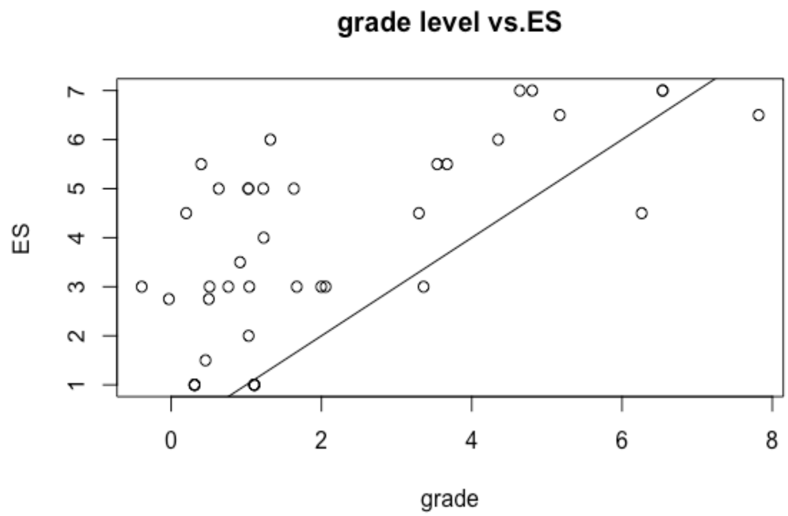


Figure 2.

The impact of grade level on the effect size of whole number computation (WNC).



Chapter 11 : Reference

Berkeley, S., Scruggs, T. E., & Mastropieri, M. A. (2010). Reading comprehension instruction for students with learning disabilities, 1995—2006: A meta-analysis. *Remedial and Special Education*, *31*(6), 423-436.

Bos, S. E., Powell, S. R., Maddox, S. A., & Doabler, C. T. (2023). A synthesis of the conceptualization and measurement of implementation fidelity in mathematics intervention research. *Journal of Learning Disabilities*, *56*(2), 95-115.

Bouck, E. C., Anderson, R. D., Long, H., & Sprick, J. (2022). Manipulative-based instructional sequences in mathematics for students with disabilities. *Teaching Exceptional Children*, *54*(3), 178-190.

Bouck, E. C., & Flanagan, S. M. (2010). Virtual manipulatives: What they are and how teachers can use them. *Intervention in School and Clinic*, *45*(3), 186-191.

Bouck, E. C., & Park, J. (2018). A systematic review of the literature on mathematics manipulatives to support students with disabilities. *Education and Treatment of Children*, *41*(1), 65-106.

Bouck, E. C., Working, C., & Bone, E. (2018). Manipulative apps to support students with disabilities in mathematics. *Intervention in School and Clinic*, *53*(3), 177-182.

Bruner, J. S. (1964). The course of cognitive growth. *American psychologist*, *19*(1), 1.

Busse, R., Kratochwill, T. R., & Elliott, S. N. (1995). Meta-analysis for single-case consultation outcomes: Applications to research and practice. *Journal of School Psychology*, *33*(4), 269-285.

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, *105*(2), 380.

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for research in Mathematics Education*, *15*(3), 179-202.

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

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R., Tankersley, M., & Test, D. W. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education*, *36*(4), 220-234.

Dennis, M. S., Sharp, E., Chovanes, J., Thomas, A., Burns, R. M., Custer, B., & Park, J. (2016). A meta-analysis of empirical research on teaching students with mathematics learning difficulties. *Learning Disabilities Research & Practice*, *31*(3), 156-168.

Domino, J. (2010). *The effects of physical manipulatives on achievement in mathematics in grades K–6: A meta-analysis*. State University of New York at Buffalo.

Flores, M. M., & Hinton, V. M. (2019). Improvement in elementary students' multiplication skills and understanding after learning through the combination of the concrete-representational-abstract sequence and strategic instruction. *Education and Treatment of Children*, *42*(1), 73-100.

Flores, M. M., Moore, A. J., & Meyer, J. M. (2020). Teaching the partial products algorithm with the concrete representational abstract sequence and the strategic instruction model. *Psychology in the Schools*, *57*(6), 946-958.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of educational research*, *79*(3), 1202-1242.

Hedges, L. V. (1994). Fixed effects models. *The handbook of research synthesis*, *285*, 299.

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

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*(1), 39-65.

Holmes, A. B. (2013). Effects of Manipulative Use on PK-12 Mathematics Achievement: A Meta-Analysis. *Society for Research on Educational Effectiveness*.

Initiative, C. C. S. S. (2010). National governors association center for best practices and council of chief state school officers. *Retrieved December*, *11*, 2012.

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

Knapp, G., & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in medicine*, *22*(17), 2693-2710.

Lafay, A., Osana, H. P., & Valat, M. (2019). Effects of interventions with manipulatives on immediate learning, maintenance, and transfer in children with mathematics learning disabilities: A systematic review. *Education Research International*, *2019*.

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

Miller, S. P., & Kaffar, B. J. (2011). Developing addition with regrouping competence among second grade students with mathematics difficulties. *Investigations in Mathematics Learning*, *4*(1), 24-49.

Moyer, P. S., Bolyard, J. J., & Spikell, M. A. (2002). What are virtual manipulatives? *Teaching children mathematics*, *8*(6), 372-377.

Moyer-Packenham, P. S., & Westenskow, A. (2013). Effects of virtual manipulatives on student achievement and mathematics learning. *International Journal of Virtual and Personal Learning Environments*, *4*(3), 35-50.

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

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Brennan, S. E. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International journal of surgery*, *88*, 105906.

Peltier, C., Morin, K. L., Bouck, E. C., Lingo, M. E., Pulos, J. M., Scheffler, F. A., . . . Deardorff, M. E. (2020). A meta-analysis of single-case research using mathematics manipulatives with students at risk or identified with a disability. *The Journal of Special Education*, *54*(1), 3-15.

Piaget, J. (1962). Szimbólumképzés a gyermekkorban [Play, dreams and imitation in childhood. New York, NY. In: Norton.

Rohatgi, A. (2014). WebPlotDigitizer user manual version 3.4. *URL http://arohatgi. info/WebPlotDigitizer/app*, 1-18.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological bulletin*, *86*(3), 638.

Rosner, B. (1975). On the detection of many outliers. *Technometrics*, *17*(2), 221-227.

Rosnow, R. L., & Rosenthal, R. (1992). Statistical procedures and the justification of knowledge in psychological science.

Shin, M., Park, J., Grimes, R., & Bryant, D. P. (2021). Effects of using virtual manipulatives for students with disabilities: Three-level multilevel modeling for single-case data. *Exceptional Children*, *87*(4), 418-437.

Zelinsky, N. A., & Shadish, W. (2018). A demonstration of how to do a meta-analysis that combines single-case designs with between-groups experiments: The effects of choice making on challenging behaviors performed by people with disabilities. *Developmental Neurorehabilitation*, *21*(4), 266-278.