Hammill Institute on Disabilities

The Journal of Special Education
2020, Vol. 54(1) 3–15

DOI: 10.1177/0022466919844516
journalofspecialeducation.sagepub.com
SAGE

A Meta-Analysis of Single-Case Research Using Mathematics Manipulatives With Students At Risk or Identified With a Disability

Corey Peltier, PhD[1], Kristi L. Morin, PhD[2], Emily C. Bouck, PhD[3], Mindy E. Lingo, MEd[1], Joshua M. Pulos, MEd[1], Faye A. Scheffler, MEd[1], Andrea Suk, MSEd[1], Leslie A. Mathews, MEd[1], Tracy E. Sinclair, MEd[1], and Malarie E. Deardorff, MEd[1]

Abstract

Manipulatives are widely considered an effective practice and have been recommended as an evidence-based practice for students identified with a learning disability when used within the concrete–representational–abstract instructional framework. The aim of the current study was to evaluate single-case experimental designs that implemented a mathematics intervention using manipulatives on the mathematical outcomes of students at risk or identified with a disability. A total of 53 studies were included in the review. The Tau-*U* effect size (ES) across studies ranged from 0.34 to 1.00, with an omnibus ES of 0.91 (CI_{95} = [0.87, 0.95]). The between-case standardized mean difference for individual studies ranged from 0.03 to 18.58. Moderator analyses revealed that out of nine variables analyzed (i.e., study quality, design, age, interventionist, manipulative type, perceptual richness, math concept, dependent variable, and disability category), only disability category served as a moderator. Implications for research and practice are discussed.

Keywords

manipulatives, mathematics, meta-analysis, single case

Reading, writing, and arithmetic are the backbone of modern education for students in PK–12. The mathematics performance for students with disabilities serves as an early indicator for in-school and postschool success (Test, Mazzotti, Fowler, Kortering, & Kohler, 2009). Yet, recent data indicate students struggle with mathematics. In 2017, students without disabilities in fourth grade scored an average of 7 points below the cut score to indicate proficiency in the mathematics portion of the National Assessment of Educational Progress (NAEP, 2017); fourth graders with disabilities scored 36 points below. For eighth graders, a similar pattern emerges. Both students with (i.e., 247) and without disabilities (i.e., 288) were below the proficient cut score of 300, although the difference between the two groups was greater (NAEP, 2017). Given the large discrepancy between students' actual scores and scores indicating proficiency, it is imperative that effective mathematical instructional practices are identified that can help mitigate this gap.

One instructional practice used to support students in developing both conceptual and procedural understanding of mathematical content is manipulatives. Manipulatives are largely defined as concrete objects that students can manipulate to support learning, and their use during mathematics instruction is widely promoted by researchers in mathematics education (Moyer, 2001) and special education (Bouck & Park, 2018). Systematic reviews on the use of manipulatives to improve mathematical outcomes for students with and without disabilities report similar conclusions: Manipulatives are effective at improving mathematical outcomes (Bouck & Park, 2018; Bouck, Satsangi, & Park, 2018; Carbonneau, Marley, & Selig, 2013). The benefit of manipulatives to support students' understanding and procedural skills with mathematical concepts was found across mathematical content and grade levels for students with and without disabilities (Bouck & Park, 2018; Bouck, Satsangi, & Park, 2018; Carbonneau et al., 2013).

[1]The University of Oklahoma, Norman, USA
[2]The University of North Carolina at Chapel Hill, USA
[3]Michigan State University, East Lansing, USA

Corresponding Author:
Corey Peltier, Department of Educational Psychology, The University of Oklahoma, 820 Van Vleet Oval, Norman, OK 73019, USA.
E-mail: coreypeltier@ou.edu

From these systematic reviews, multiple points can be discerned. First, although researchers found manipulatives are effective when used alone, they are also effective when used as part of an instructional framework: the concrete–representational–abstract (CRA) framework (Bouck & Park, 2018; Bouck, Satsangi, & Park, 2018). Second, outcomes for students at risk or identified with a disability were maximized when manipulatives were paired with explicit instruction. Third, virtual manipulatives showed promising results within the CRA framework and in isolation. Fourth, despite the sustained examination of the use of manipulatives, further examination for whom, by whom, and under what conditions manipulatives are most effective is warranted. Finally, the methodological rigor of studies was evaluated with eight of 20 (Bouck, Satsangi, & Park, 2018) and nine of 36 (Bouck & Park, 2018) studies meeting indicators (i.e., Cook et al., 2015; Gersten et al., 2005; Horner et al., 2005).

Carbonneau et al. (2013), in their meta-analysis of manipulatives for students without disabilities, analyzed multiple moderators of interest. First, the mathematical concept was evaluated with fractions being statistically significantly greater than other mathematical concepts. The reported effects for place value, arithmetic, geometry, and algebra were all between 0.21 and 0.58, but were not statistically significantly different from one another. One further area of exploration was whether the perceptual richness of a manipulative moderated effects. The authors grounded their work based off prior research (Kaminski, Sloutsky, & Heckler, 2009; McNeil, Uttal, Jarvin, & Sternberg, 2009), which suggested perceptually rich (i.e., realistic) manipulatives hinder students' ability to learn new mathematical concepts because children fail to generalize the concrete object to the abstract mathematics they represent. On immediate tasks, the perceptual richness did not affect results; however, as the task became more distal (i.e., transfer, problem-solving), the bland manipulatives yielded statistically significant larger effects. Finally, Carbonneau et al. (2013) investigated whether the use of manipulatives was differential across ages of participants (i.e., 3–6, 7–11, 12 years, and older). The decision was made based on developmental theorists' work suggesting a child's developmental status would dictate how knowledge is learned (Bruner, 1964; Piaget, 1962). Children in the age group of 7 to 11 years benefited more from using manipulatives than students 12 years and older, which aligned with theorists' developmental theory. Of note, these developmental stages were developed based on typically developing children; further investigation on the effectiveness of manipulatives with students who are atypical (i.e., identified with a disability) is warranted.

Purpose of the Current Study

The aim of a meta-analysis is to aggregate and evaluate an existing body of research to investigate unanswered questions on a topic (Borenstein, Hedges, Higgins, & Rothstein, 2009). The field of special education has identified indicators to be used to evaluate the quality of research, with some (i.e., U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse [WWC], 2017) recommending the removal of studies failing to meet standards from analyses. However, a meta-analytic approach allows for the data to speak for itself by including *all* data and considering methodological quality as perhaps an indicator of the dispersion in effect sizes (ESs; Cooper, 2016).

The current study represents a meta-analytic investigation of the effects of manipulatives on students' mathematical outcomes when evaluated via a single-case experimental design (SCED). We aimed to extend the literature by evaluating the methodological quality using the WWC Design Standards and identifying whether effects vary with quality; reporting effects using visual analysis, a robust nonoverlap index (Tau-*U*; Parker, Vannest, Davis, & Sauber, 2011), and a between-case ES analogous to Cohen's *d* (between-case standardized mean difference [BC-SMD]; Pustejovsky, Hedges, & Shadish, 2014); and investigating whether effects vary based on systematic differences related to intervention design or population characteristics. Results from this review will aid practitioners in manipulative selection and intervention design and provide estimates of student outcomes based on demographic characteristics to maximize student outcomes. For researchers, the study will identify areas for systematic replication and areas of needed research. The following research questions will be addressed:

> **Research Question 1:** What is the effect of mathematics interventions on child outcomes using manipulatives, reported via visual analysis, Tau-*U*, and BC-SMD?
> **Research Question 2:** What effect does methodological quality have on the effectiveness of mathematics interventions using manipulatives?
> **Research Question 3:** What effects do participant characteristics (i.e., age, disability), manipulative characteristics (i.e., concrete or virtual, perceptually rich or bland), and interventionist have on the effectiveness of mathematics interventions using manipulatives?
> **Research Question 4:** How do effects vary across dependent measures (i.e., mathematical concept, dependent variable measurement procedure)?

Method

Study Identification

Search strategy. The following databases were searched by combining the term *manipulative** with *app** OR *computer** OR *virtual** OR *digital** OR *technolog** OR *math**

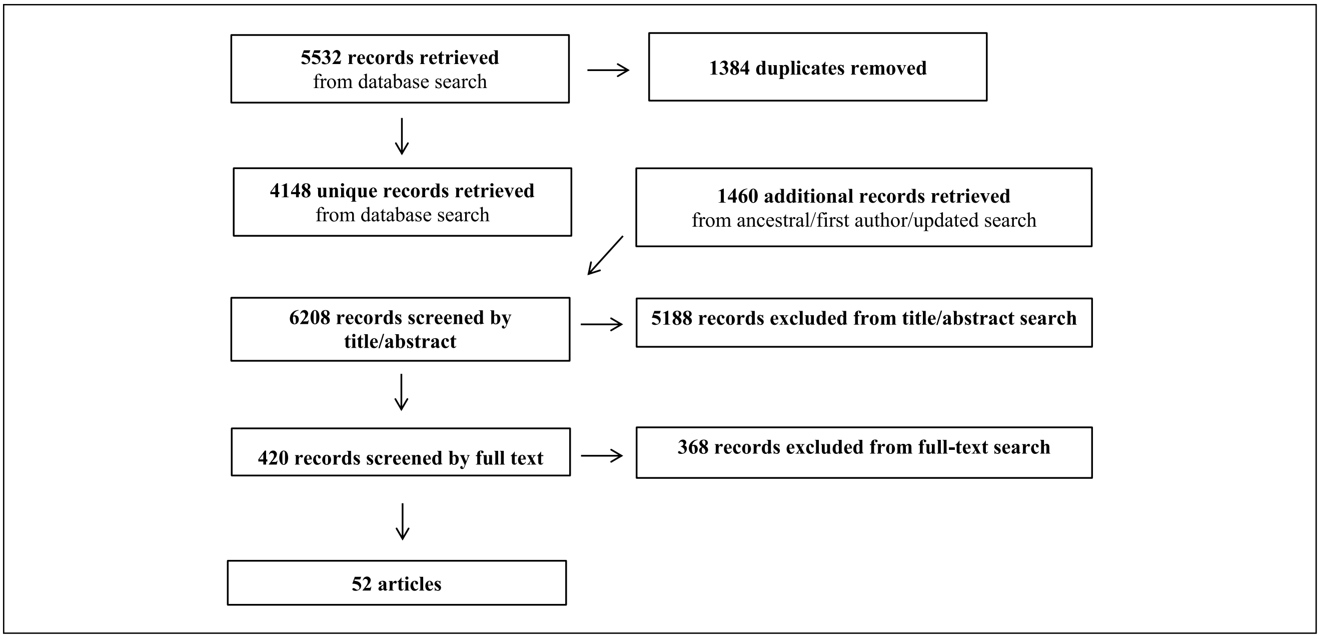

Figure 1. Flowchart depicting the records included and excluded during each phase.

OR *concrete** OR *physical**: *Education Resources Information Center* (ERIC), *PsycARTICLES, PsycINFO, Education Source*, and *Teacher Reference Center*. The search was limited to academic journals and dissertations. In addition, an ancestral, forward, and first author searches were conducted on all documents identified as meeting the inclusion criteria. Article title or first author name was entered into SCOPUS to identify potential references. If an article was not available in SCOPUS, a hand search of the reference list was conducted. Finally, a hand search of the curriculum vitae of researchers known to research math manipulatives was conducted.

Inclusion criteria. To be included, studies needed to meet the following inclusion criteria: (a) published in English, (b) peer-reviewed article or dissertation, (c) used an SCED, (d) primary intervention component was the use of a manipulative (i.e., a concrete or virtual/digital object a student would manipulate or move to aid in understanding or solving mathematics problems), (e) included at least one student outcome related to mathematics, (f) implemented in a school setting (i.e., pre-kindergarten through 12th grade), and (g) included participants at risk or identified with a disability. The decision to include students identified as at risk was made because (a) students at risk displayed similar performance levels as students identified with a learning disability on universal screener data and (b) within a multitiered system of support framework, students at risk would be receiving secondary or tertiary instruction alongside students identified with a disability.

Title/abstract and full-text review. After removing duplicates, the systematic search resulted in 4,148 articles. Titles and abstracts of all identified articles were evaluated against the inclusion criteria. If a decision could not be made on the title/abstract alone, the article was retained for full-text screening. Next, full-text screening against the inclusion/exclusion criteria was conducted. An ancestral, first author, and forward searches were conducted on the included articles (see Figure 1 for the full process).

Methodological Evaluation

The methodological quality of studies was evaluated using the WWC Pilot Single-Case Design Standards (Kratochwill et al., 2010; U.S. Department of Education, Institute of Education Sciences, WWC, 2017) to determine whether (a) the independent variable was systematically manipulated, (b) interobserver agreement data were reported, (c) interobserver agreement data were collected on a minimum of 20% of data in both baseline and intervention phases, (d) interobserver agreement scores met minimum quality thresholds (i.e., 80% or higher for percent agreement or .60 or higher for kappa), (e) there were a minimum of three attempts to demonstrate treatment effects at three different points in time, and (f) there were a minimum of three data points in baseline and intervention phases for multiple-baseline and multiple-probe experiments or a minimum of four data points for alternating treatment designs. Studies were evaluated at the experimental level, meaning each experimental design (i.e., ABAB, multiple-baseline design) was evaluated.

Variable Coding

A coding menu was created in Qualtrics by referring to the work conducted by previous researchers (Bouck & Park, 2018; Carbonneau et al., 2013). Each study was coded for the following information: (a) publication type, (b) type of SCED, (c) interventionist, (d) description of training provided to interventionist, (e) demographic/experience of interventionist, (f) demographic characteristics of participants, (g) type of manipulatives used (i.e., concrete, virtual), (h) description of manipulative(s) used to determine whether it was perceptually rich or bland, (i) paradigm of instruction (i.e., explicit, constructivist), (j) use of the CRA framework or some other variation, (k) duration of the intervention, (l) behavioral component to the intervention (e.g., token economy), (m) academic/instructional component, (n) description of the fidelity checklist/procedures used, (o) description of baseline prior to intervention, (p) dependent measure(s), (q) mathematical concepts, (r) generalization, and (s) maintenance collected.

Visual Analysis

Because SCEDs have historically been evaluated using visual analysis, the authors conducted a visual analysis, in addition to the statistical analysis, of all experiments. A researcher-developed coding sheet aligned with the WWC Procedures and Standards Handbook Studies (U.S. Department of Education, Institute of Education Sciences, WWC, 2017) was used to evaluate the effectiveness of the intervention (see Morin et al., 2018). The following characteristics were evaluated: (a) level, (b) trend, (c) variability, (d) immediacy of effect, (e) proportion of overlap, and (f) consistency of data across phases. The trend, level, and variability were analyzed in each phase independently. The immediacy of effect, proportion of overlap, and consistency of data were analyzed by comparing adjacent phases (i.e., baseline and intervention). Each experiment was identified as reporting strong, moderate, or weak evidence of effect.

Data Extraction and ES Calculation

Data extraction. GraphClick (Arizona Software, 2010) was used to extract data from each graph because it has been shown to have high reliability (Boyle, Samaha, Rodewald, & Hoffmann, 2013). A JPEG of each experiment was imported into the program. The ordinate was set to scale and the data points were plotted. The program provided the digitized results and they were exported to an Excel file.

Data analysis. Intervention effects were reported by calculating more than one ES as recommended by the field (Manolov, Guilera, & Solanas, 2017; Vannest, Peltier, & Haas, 2018). This decision was made because there is no consensus on one ES to use for SCED data (Kratochwill et al., 2013; Pustejovsky, 2018) and the practice of reporting multiple ESs has grown (Losinski, Ennis, Sanders, & Nelson, 2018; Maggin, Pustejovsky, & Johnson, 2017).

The field has raised concerns regarding the use of meta-analytic techniques with SCEDs (Baron & Derenne, 2000; Burns, 2012). Concerns arise with quantifying effects using group design metrics (e.g., Cohen's *d*) that do not fit data (Burns, 2012), and ES metrics weaken causal claims and the idiosyncrasies of the data are lost (Baron & Derenne, 2000; Salzberg, Strain, & Baer, 1987). Furthermore, ES metrics commonly used with SCED data are influenced by outliers and do not have the ability to report confidence intervals (CIs; Busk & Serlin, 1992; Riley-Tillman & Burns, 2009).

Recent advances in the statistical analysis of SCEDs led to the creation of a nonoverlap ES (i.e., Tau-*U*) that allows the user to adjust for undesirable baseline and report CIs (Parker, Vannest, & Davis, 2011). Furthermore, Tau-*U* has been shown to be more robust than other nonoverlap ESs (Parker et al., 2011) and yield high correspondence with visual analysis (Brossart, Vannest, Davis, & Patience, 2014). However, a major limitation to nonoverlap ESs is the inability to account for the magnitude. Hedges, Pustejovsky, and Shadish (2012, 2013; Shadish, Hedges, & Pustejovsky, 2014) introduced BC-SMD that is analogous to Cohen's *d*; thus, it takes magnitude of change into account and is comparable with group design ESs.

To calculate Tau-*U*, data from each baseline (Phase A) and intervention (Phase B) were entered into a web-based Tau-*U* calculator (http://www.singlecaseresearch.org/calculators/tau-u; Vannest, Parker, Gonen, & Adiguzel, 2016). If a baseline had an undesirable trend (i.e., a trend in a therapeutic direction), the ES was corrected by selecting this option. Next, each Tau-*U* ES and its standard error (SD_{Tau}) per experiment-level AB phase contrast was entered into the Comprehensive Meta-Analysis software program (Version 3; Borenstein, Hedges, Higgins, & Rothstein, 2005) to calculate a study-level Tau-*U*, an omnibus effect Tau-*U*, and to conduct moderator analyses. An inverse variance weighting scheme was used.

We used a web-based calculator (https://jepusto.shinyapps.io/scdhlm/; Pustejovsky, 2016) to calculate the BC-SMD. To ensure data were read correctly, the authors used a data template provided on the website. For reversal and multiple-baseline designs, we opted to use restricted maximum likelihood (REML) estimation because it is a more flexible model. The current model specifies a fixed effect and random effect for baseline level be used because it is unlikely to assume the average outcome is zero across cases or the same. For the intervention level, we selected a fixed effect for all cases, signifying a change in level from baseline to intervention, and a random effect for all cases, signifying differences in intervention effect across cases.

For multiple-baseline designs across participants, we specified the model to allow for fixed effects at baseline because data were greater than zero but typically consistent across participants. A fixed-effect and random effects model were specified for intervention phase because upon visual inspection we identified most experiments demonstrated a change in level from baseline to intervention phase and this intervention effect was differential across cases. Each model was specified by identifying whether there were phase time trends (i.e., level, linear) during baseline and/or intervention data.

We hypothesized variance in study effects was due to systematic differences rather than sampling error alone; thus, we specified that a random effects model was preferable (Borenstein, Hedges, Higgins, & Rothstein, 2009). The Q statistic (Cochran, 1954) and I^2 index (Higgins & Thompson, 2002) were reported to identify whether there was a statistically significant amount of variance between reported ESs. To conduct moderator analyses for nominally scaled variables, the total variance identified (i.e., Q_{total}) was partitioned into the variance within each subgroup (i.e., Q_{within}) and variance between the groups (i.e., Q_{between}). The Q_{between} quantified the amount of variance explained by that moderator variable. Analogous to interpreting analyses of variance (ANOVAs), we interpreted the Q_{within} to identify whether the homogeneity of variance assumption was violated, meaning each subgroup retained high levels of variance. We interpreted the Q_{between} to identify whether the variable explained a statistically significant amount of the variance in ESs, thus functioning as a moderator.

Publication bias. We computed *Rosenthal's fail-safe N* (Begg & Mazumdar, 1994; Rosenthal, 1979) to identify the potential of publication bias in the studies included in this meta-analysis. The *Rosenthal's fail-safe N* provides an estimated number of nonsignificant studies that would need to be added to "nullify" the reported omnibus effect. To interpret *Rosenthal's fail-safe N*, we used the suggested criterion to obtain an estimate of the desired *fail-safe N*: 5 × (studies included in review) + 10 (Becker, 2005; Rosnow & Rosenthal, 1989). In the current study, we calculated $5 \times 48 + 10 = 250$. The obtained *fail-safe N* ($n = 2{,}066$) was larger than the desired *fail-safe N*, which we interpreted to mean publication bias is not a concern.

Interrater reliability. Interrater reliability (IRR) was conducted on a minimum of 30% of all titles/abstract screenings, 37% of all full-text screenings, 100% of the WWC Design Standard coding, 25% of variable coding, 65% of visual analysis, and 11.5% of data extraction. If any IRR disagreement occurred, the primary and secondary raters met to discuss the disagreement until a consensus was reached. IRR results were as follows: 95% for titles/abstract screening, 83% for full-text screening, 84% for the WWC Design Standard coding, 95% for study characteristics, and 83% for visual analysis. Disagreements for the WWC Design Standard coding primarily occurred over design standards 2B (interobserver agreement frequency) and 2C (interobserver agreement quality) due to the ambiguous wording used to describe interobserver agreement. For study characteristics, the primary variable of disagreement was in regard to the type of manipulative used (rich or bland) and the type of mathematical task. For visual analysis, the primary disagreement was in regard to determining whether three intervention effects were demonstrated. Disagreements on this criterion affected agreement on the subsequent criteria.

Results

Overall Effects

Studies meeting the inclusion criteria included students identified with various disabilities and targeted diverse mathematical concepts (see Supplemental Appendices A and B). A total of 53 studies met the inclusion criteria (see Supplemental Appendix E for reference list); 48 studies (335 AB phase contrasts, $n = 211$) were included in the omnibus ES (reasons for exclusion: alternating treatment design [$k = 3$], AB design [$k = 1$], and multiple-probe design across two [$k = 1$]). The Tau-*U* ES for studies ranged from 0.34 to 1.00, with an omnibus ES of 0.91 ($CI_{95} = [0.87, 0.95]$); see Table 1). The Q_{total} was 44.78 ($p = .57$) and the I^2 index was 0. The overall BC-SMD for individual studies ranged from 0.03 to 18.58 (see Table 2). Visual analysis revealed that 20 (20%) experiments demonstrated strong evidence, 36 (36%) experiments demonstrated moderate evidence, and 43 (43%) experiments demonstrated weak evidence (see Supplemental Appendix C).

Methodological Evaluation

A central research question was an investigation of the relationship between methodological quality (i.e., measured via WWC Design Standards) and intervention effects. A total of 11 studies (70 AB phase contrasts) met indicators without reservations, 23 studies (138 AB phase contrasts) met indicators with reservations, and 15 studies (123 AB phase contrasts) failed to meet indicators (see Supplemental Appendix D). Study quality did not moderate intervention effects. The type of SCED was also evaluated; it did not function as a moderator (see Table 3).

Moderator Analyses: For Whom and By Whom

To identify whether effects varied across participants, two variables were investigated: age and disability. Age did not function as a moderator. Interventions were less effective for students identified with an emotional or behavioral disorder than all other disability categories, except developmental delay (see Table 3). The interventionist of each intervention

Table 1. Tau-*U* Effects per Study and Omnibus.

Study	Pairs	ABs	Tau-*U*	CI Lower	CI Upper	*p* value
Agrawal (2013, Study 1)	708	12	0.96	0.79	1.00	<.01
Agrawal (2013, Study 2)	708	12	0.89	0.71	1.00	<.01
Bouck, Bassette, et al. (2017)	175	3	1.00	0.64	1.00	<.01
Bouck, Chamberlain, and Park (2017)	269	4	1.00	0.71	1.00	<.01
Bouck, Park, and Nickell (2017)	172	4	0.99	0.66	1.00	<.01
Bouck et al. (2018)	307	6	1.00	0.73	1.00	<.01
Browder, Jiminez, Spooner, et al. (2012)	920	16	0.84	0.68	1.00	<.01
Browder et al. (2018)	69	8	1.00	0.58	1.00	<.01
Burns et al. (2015)	53	2	0.34	−0.20	0.87	.22
Cass et al. (2003)	196	6	0.81	0.52	1.00	<.01
Cihak and Grim (2008)	946	12	0.99	0.82	1.00	<.01
Denny and Test (1995)	128	3	0.47	−0.17	1.00	.15
Flores (2009)	498	6	1.00	0.77	1.00	<.01
Flores (2010)	320	6	1.00	0.74	1.00	<.01
Flores, Hinton, and Schweck (2014)	547	10	0.95	0.76	1.00	<.01
Flores, Hinton, and Strozier (2014)	222	4	0.65	0.35	0.96	<.01
Flores, Hinton, and Burton (2016)	355	3	0.94	0.64	1.00	<.01
Hardy (2014)	504	8	0.95	0.72	1.00	<.01
Harris et al. (1995)	658	13	0.85	0.67	1.00	<.01
Hinton et al. (2016)	450	12	0.90	0.70	1.00	<.01
Hord and Xin (2015)	18	3	1.00	0.35	1.00	<.01
Hudson et al. (2016)	41	3	1.00	0.48	1.00	<.01
Huntington (1994)	72	6	0.39	0	0.78	.05
Jiminez et al. (2008)	253	3	0.96	0.62	1.00	<.01
Jiminez and Kemmery (2013)	106	5	0.95	0.59	1.00	<.01
Jiminez and Staples (2015)	153	3	0.85	0.48	1.00	<.01
Maccini and Ruhl (2000)	48	6	0.96	0.53	1.00	<.01
Maccini and Hughes (2000)	388	42	0.92	0.76	1.00	<.01
Mancl (2012)	204	3	0.94	0.63	1.00	<.01
Mancl et al. (2012)	241	5	0.99	0.74	1.00	<.01
Marsh and Cooke (1996)	184	3	1.00	0.65	1.00	<.01
Miller and Mercer (1993)	329	9	0.60	0.33	0.87	<.01
Morin and Miller (1998)	142	3	1.00	0.61	1.00	<.01
Ok and Bryant (2016)	322	4	0.98	0.69	1.00	<.01
Ozdemire (2018)	72	3	1.00	0.54	1.00	<.01
Reneau (2013)	84	5	0.42	0.03	0.82	.04
Root et al. (2017)	75	6	1.00	0.62	1.00	<.01
Satsangi and Bouck (2015)	236	6	1.00	0.73	1.00	<.01
Satsangi, Hammer, and Evmenova (2018)	95	3	1.00	0.59	1.00	<.01
Satsangi, Hammer, and Hogan (2018)	100	3	1.00	0.60	1.00	<.01
Saunders (2014)	207	3	1.00	0.66	1.00	<.01
Scheuermann et al. (2009)	180	14	0.95	0.71	1.00	<.01
Sealander et al. (2012)	174	8	0.42	0.07	0.77	.02
Shin and Bryant (2017)	93	3	0.73	0.29	1.00	<.01
Strickland and Maccini (2013a)	24	3	0.94	0.31	1.00	<.01
Strickland and Maccini (2013b)	48	5	1.00	0.56	1.00	<.01
Stroizer et al. (2015)	572	9	0.97	0.76	1.00	<.01
Yakubova et al. (2016)	930	12	0.97	0.81	1.00	<.01
Omnibus	13,624	335	0.91	0.87	0.95	<.01

(continued)

Table 1. (continued)

Study	Pairs	ABs	Tau-*U*	CI		*p* value
				Lower	Upper	
Tau-*U* effects for alternating treatment designs and AB designs						
Bouck et al. (2014)	360	12	0.98	0.80	1.00	<.01
Bouck, Chamberlain, and Park (2017)	170	6	0.91	0.37	1.00	<.01
Browder, Jimenez, and Trela (2012)	313	7	0.80	0.21	1.00	.02
Mulcahy and Krezmien (2009)	46	4	1.00	0.61	1.00	<.01
Satsangi et al. (2016)	300	6	1.00	0.78	1.00	<.01
Bouck, Chamberlain, and Park (2017)	170	6	0.91	0.37	1.00	<.01

Note. References for the studies included in the meta-analysis are available in the supplemental material. ABs refer to the number of AB phase contrasts. CI = confidence interval.

Table 2. BC-SMD Effects per Study.

Study	BC-SMD	*SE*	CI		*df*	Auto	ICC	Initial	Follow-up
			Lower	Upper					
Agrawal (2013, Study 1)	2.14	0.44	1.32	3.03	34.72	0.55	0.01	5	16
Agrawal (2013, Study 1)	2.10	0.69	0.90	3.52	12.28	0.86	0.12	5	16
Agrawal (2013, Study 2)	1.36	0.53	0.44	2.43	11.09	0.87	0.19	5	16
Agrawal (2013, Study 2)	0.57	0.38	−0.10	1.32	9.51	0.93	0.02	5	16
Bouck, Bassette, et al. (2017)	2.60	1.05	1.13	4.82	6.04	0.21	0.40	3	16
Bouck, Park, and Nickell (2017)	6.11	0.66	4.92	7.47	51.85	−0.05	0.00	5	14
Bouck, Park, et al. (2017)	3.79	0.71	2.51	5.25	27.71	0.27	0.01	3	10
Browder et al. (2018)	7.52	1.99	4.46	11.77	10.22	−0.66	0.82	5	7
Cass et al. (2003)	0.46	0.43	−0.25	1.29	5.53	0.78	0.00	5	11
Cass et al. (2003)	0.07	0.27	−0.35	0.52	3.29	0.93	0.00	3	10
Denny and Test (1995)	0.51	0.75	−0.71	1.91	4.23	0.87	0.01	4	9
Flores (2009)	0.73	0.24	0.34	1.23	9.38	0.95	0.00	3	18
Flores (2010)	0.49	0.40	0.03	1.24	3.33	0.94	0.00	3	9
Flores (2010)	1.31	0.86	0.44	2.99	3.45	0.95	0.00	3	16
Flores, Hinton, and Strozier (2014)	0.03	0.19	−0.31	0.04	5.54	0.94	0.00	5	15
Flores, Hinton, and Burton (2016)	2.58	0.51	1.67	3.62	25.96	0.40	0.00	5	22
Harris et al. (1995)	0.86	0.21	0.47	1.27	42.19	0.59	0.15	2	9
Hord and Xin (2015)	2.94	1.01	1.42	5.07	7.70	−0.53	0.32	3	5
Hudson et al. (2016)	1.13	0.78	0.26	2.65	3.52	0.86	0.00	3	8
Huntington (1994)	1.56	1.17	−0.47	3.87	6.94	−0.88	0.21	4	7
Huntington (1994)	9.05	9.92	−9.65	28.44	14.62	−0.01	0.00	4	7
Jimenez et al. (2008)	0.20	0.22	−0.09	0.58	3.22	0.97	0.00	3	29
Jimenez and Kemmery (2013)	0.77	0.39	0.23	1.57	5.13	0.44	0.82	5	9
Jimenez and Staples (2015)	1.23	0.55	0.38	2.37	6.54	0.38	0.32	5	15
Maccini and Ruhl (2000)	2.06	0.74	0.81	3.59	10.70	0.17	0.00	4	6
Maccini and Ruhl (2000)	2.94	1.60	0.79	6.20	4.65	0.18	0.47	4	6
Maccini and Hughes (2000)	4.48	2.31	1.82	9.24	4.23	−0.47	0.64	4	6
Maccini and Hughes (2000)	1.00	0.96	−0.16	2.74	3.19	−0.42	0.74	4	6
Maccini and Hughes (2000)	1.08	1.81	−0.08	3.42	2.29	−0.27	0.92	4	6
Maccini and Hughes (2000)	1.34	1.14	−0.07	3.49	3.41	−0.34	0.73	4	6
Maccini and Hughes (2000)	2.01	0.83	0.68	3.73	7.56	0.04	0.32	4	6
Maccini and Hughes (2000)	1.65	0.78	0.42	3.26	6.61	−0.21	0.43	4	6
Maccini and Hughes (2000)	1.39	0.64	0.27	2.68	10.50	0.31	0.00	4	6
Maccini and Hughes (2000)	2.38	0.62	1.30	3.67	15.13	−0.31	0.00	4	6

(continued)

Table 2. (continued)

Study	BC-SMD	*SE*	CI Lower	CI Upper	*df*	Auto	ICC	Initial	Follow-up
Maccini and Hughes (2000)	2.36	1.35	0.73	5.08	4.10	0.46	0.46	4	6
Maccini and Hughes (2000)	1.42	0.70	0.22	2.84	8.86	0.31	0.10	4	6
Mancl (2012)	1.23	1.19	−0.01	3.35	2.95	−0.11	0.77	3	20
Mancl et al. (2012)	6.97	0.75	5.60	8.52	53.32	−0.28	0.00	3	14
Marsh and Cooke (1996)	3.39	1.80	1.34	7.08	4.13	0.55	0.56	5	17
Miller and Mercer (1993)	1.72	0.79	0.36	3.33	9.36	0.40	0.22	3	7
Miller and Mercer (1993)	0.88	0.90	−0.84	2.64	27.12	−0.47	0.00	5	10
Miller and Mercer (1993)	1.20	0.56	0.34	2.35	6.51	0.28	0.40	3	6
Morin and Miller (1998)	2.52	1.16	1.07	4.94	4.90	0.10	0.48	3	14
Ok and Bryant (2016)	1.47	0.46	0.67	2.42	13.33	0.67	0.02	3	18
Ozdemire (2018)	6.13	4.28	2.21	14.36	3.23	0.00	0.75	5	8
Reneau (2013)	1.08	0.37	0.45	1.84	11.68	0.01	0.46	4	7
Root et al. (2017)	2.58	2.76	0.68	7.07	2.57	0.29	0.77	2	6
Root et al. (2017)	2.19	3.65	0.25	6.77	2.26	0.66	0.83	2	5
Satsangi and Bouck (2015)	9.36	2.03	6.02	13.65	15.01	0.34	0.00	5	10
Satsangi and Bouck (2015)	5.85	1.04	3.99	7.99	28.16	−0.11	0.00	5	10
Satsangi, Hammer, and Evmenova (2018)	18.58	7.86	9.01	42.45	5.05	0.07	0.54	5	10
Satsangi, Hammer, and Hogan (2018)	10.87	2.19	7.30	15.54	15.53	0.33	0.03	5	11
Saunders (2014)	4.27	0.92	2.67	6.17	19.72	0.38	0.03	5	14
Scheuermann et al. (2009)	3.46	0.40	2.72	4.29	52.37	0.27	0.16	3	7
Sealander et al. (2012)	0.36	0.28	−0.18	0.91	28.01	0.32	0.20	3	8
Shin and Bryant (2017)	1.47	0.79	0.09	3.06	9.13	0.26	0.06	3	12
Strickland and Maccini (2013a)	12.38	3.20	7.54	19.26	10.05	0.05	0.10	2	4
Strickland and Maccini (2013b)	8.06	3.27	3.90	14.99	5.41	−0.03	0.81	2	5

Note. References for the studies included in the meta-analysis are available in the supplemental material. BC-SMD = between-case standardized mean difference; CI = confidence interval; auto = autocorrelation; ICC = intraclass correlation.

was investigated to identify whether effects varied across implementers; it did not function as a moderator.

Moderator Analyses: Under What Conditions

The following variables were investigated in regard to instructional conditions: manipulative format, manipulative type, measurement creation, and nature of dependent variable. Neither manipulative format nor type functioned as a moderator. Neither measurement creation of the nature of the dependent variable functioned as a moderator.

Discussion

The primary aim of the current systematic review and meta-analysis was to evaluate the evidence from single-case research to support the use of manipulatives during mathematics instruction. Using manipulatives during mathematics instruction is recommended as best practice in general education (e.g., The National Mathematics Advisory Panel, 2007), educational psychology (e.g., Carbonneau et al., 2013), and special education (e.g., Bouck & Park, 2018). Of the 53 studies found that used an SCED to explore the use of manipulatives to support mathematics teaching and learning, only 48 studies were included in the omnibus ES due to limitation with aggregating data from alternating treatment designs and nonexperimental designs (i.e., AB). The overall ES suggested manipulatives were effective at improving the mathematical performance of students identified or at risk of being identified with a disability.

Effects were not differential across design quality. This dictated future analytic decisions; the authors opted to retain all studies in moderator analyses rather than running with and without low-quality studies. Of the 15 studies failing to meet indicators, nine were published after the special issue in *Exceptional Children* that provided recommendations for researchers on characteristics to consider and report on for single-case research (Horner et al., 2005). Researchers are encouraged to consult indicators for methodological quality when designing experiments. SCEDs allow researchers to determine whether an independent variable (i.e., intervention) caused change in the dependent variable (i.e., behavior). However, studies that fail to meet quality indicators have an increased likelihood of their internal validity being threatened, which potentially turns this causal design into correlational.

Effects regarding the effectiveness of manipulatives for supporting mathematics instruction were consistent across

Table 3. Effect Sizes by Potential Moderator.

Variable	ABs	Tau-*U* (95% CI)	Q_b	*p* Value
Methodology				
WWC rating			3.98	.14
Does not meet	123	0.86 [0.78, 0.94]		
Meets with reservations	138	0.89 [0.83, 0.94]		
Meets	70	0.96 [0.89, 1.00]		
SCED design			7.27	.06
MBD-B	9	0.97 [0.76, 1.00]		
MBD-P	82	0.82 [0.74, 0.89]		
MPD-B	76	0.94 [0.87, 1.00]		
MPD-P	164	0.93 [0.87, 0.99]		
For whom				
Age			0.82	.84
Preschool	23	0.94 [0.80, 1.00]		
Elementary	155	0.89 [0.83, 0.94]		
Intermediate	50	0.93 [0.83, 1.00]		
Secondary	103	0.90 [0.82, 0.98]		
Disability			14.50	.03
At risk	45	0.91 [0.81, 1.00]		
ASD	68	0.96 [0.88, 1.00]		
DD	6	0.91 [0.63, 1.00]		
EBD	7	0.46 [0.20, 0.71]		
ID	57	0.91 [0.81, 1.00]		
OHI	4	0.74 [0.37, 1.00]		
SLD	144	0.88 [0.82, 0.95]		
By whom				
Interventionist			1.09	.30
Researcher	205	0.92 [0.87, 0.97]		
Teacher	126	0.88 [0.81, 0.94]		
Under what conditions				
Manipulative type			2.18	.14
Concrete	294	0.89 [0.85, 0.93]		
Virtual	37	0.98 [0.87, 1.00]		
Perceptual richness			1.04	.59
Bland	223	0.93 [0.88, 0.97]		
Rich	45	0.88 [0.78, 0.98]		
Bland, rich	55	0.88 [0.79, 0.98]		
Math concept			2.46	.93
Algebra	21	0.92 [0.76, 1.00]		
Basic facts	49	0.87 [0.77, 0.96]		
Computation	34	0.94 [0.83, 1.00]		
Early numeracy	32	0.90 [0.78, 1.00]		
Fractions	31	0.94 [0.83, 1.00]		
Geometry	24	0.93 [0.78, 1.00]		
Money	26	0.85 [0.73, 0.96]		
Problem-solving	109	0.89 [0.81, 0.98]		
Dependent variable			0.92	.34
Process	103	0.93 [0.86, 1.00]		
Solution	228	0.89 [0.85, 0.93]		

Note. ABs refer to the number of AB phase contrasts. CI = confidence interval; WWC = What Works Clearinghouse; SCED = single-case experimental design; MBD-B = multiple-baseline design across behaviors; MBD-P = multiple-baseline design across participants; MPD-B = multiple-probe design across behaviors; MPD-P = multiple-probe design across participants; ASD = autism spectrum disorder; DD = developmental delay; EBD = emotional and behavioral disorder; ID = intellectual disability; OHI = other health impairment; SLD = specific learning disability.

the age of participants included, contrary to findings from Carbonneau et al. (2013), who reported lesser effects for older (i.e., secondary) than younger (i.e., elementary) students. This finding could be due to (a) differences in coding (i.e., school age coding in this study vs. Piaget's stages of development in the Carbonneau et al. (2013) study) or (b) Piaget's stages of development that are based on a typically developing child, and the population of children included in the study were identified with a disability or at risk of being identified with a disability. The current findings suggest teachers working with students identified with a disability or at risk of mathematical failure should incorporate manipulatives regardless of student age.

An additional finding related to population characteristics was manipulatives were less effective for students identified with an emotional or behavioral disorder than students identified as at risk, with autism spectrum disorder, with an intellectual disability, or a learning disability. This finding should be interpreted with caution due to the limited sample (seven students identified with emotional and behavioral disorder [EBD]); however, it may raise concerns with providing a strictly academic intervention without incorporating a behavioral component for this population. Incorporating a behavioral component (e.g., self-management, token economy) in addition to an academic intervention using manipulatives may yield improved effects in both research and practice (Losinski et al., 2018). In the studies to date, six included a behavioral component in addition to an academic component, only one study included students with EBD.

Effects were also consistent across implementer (i.e., teacher vs. researcher), which is consistent with findings from Carbonneau et al. (2013). This is promising because findings suggest, with training, teachers can implement the intervention and yield similar effects as researchers with expertise in the intervention. However, two areas the authors aimed to investigate a priori but were unable to investigate due to insufficient information in the included articles were the roles treatment fidelity and training have on treatment effects. Many authors of the included studies did not provide sufficient information on the type of training provided, duration of training, and whether coaching was provided during the implementation of the intervention. Furthermore, the thoroughness of fidelity checklists varied throughout the articles. Reasons for this lack of information are likely varied and influenced by page limit requirements; however, future researchers should aim to provide a replicable description of training and coaching procedures along with conducting a thorough task analysis of behaviors teachers will exhibit during the instructional sequence, even if this information must be included as appendices and stored in digital repositories to save journal space.

Effects were consistent with prior meta-analyses and systematic reviews across two variables related to the manipulatives used: (a) concrete or virtual and (b) perceptually rich or bland. More recently, researchers investigated the potential of using virtual manipulatives in mathematics instruction for students with and without disabilities. The premise of using and examining virtual manipulatives is their ease of use (i.e., access multiple types in a couple clicks) and their potential to represent a larger breadth of mathematical concepts (i.e., linear algebra) with more accuracy than concrete manipulatives (Sarama & Clements, 2016). Findings to date suggest virtual manipulatives are effective for students, although additional research is needed and still emerging (Bouck & Park, 2018).

In terms of bland versus perceptually rich manipulatives, the findings mirror Carbonneau et al. (2013); however, Carbonneau et al. (2013) reported bland manipulatives yielded larger effects for maintenance, problem-solving, and generalized (i.e., transfer) tasks. Given the current data, in this study, we were unable to analyze these relationships due to statistical techniques not being appropriate and the limited number of studies that reported these data.

Finally, effects were consistent across two variables related to the dependent measure: (a) mathematical concept (e.g., basic operations, fractions, and algebra) and (b) dependent variable (i.e., process vs. solution). The findings differ from Carbonneau et al. (2013), who reported concrete manipulatives were more effective for fractions than arithmetic and algebra. Although this result was not found for students identified or at risk of being identified for a disability, researchers do find that selecting an appropriate manipulative (i.e., concrete as opposed to virtual as well as the specific type of manipulative) for a given mathematical concept is important (Bouck & Park, 2018). Another area investigated was whether manipulatives yielded differential effects depending on whether students' work was evaluated based on their mathematical process (i.e., process to solve the problem) versus solution alone (i.e., right/wrong). Results from this analysis suggest manipulatives were equally effective at improving students' mathematical process and solution, which is beneficial because both are important for mathematical achievement.

Limitations and Implications for Research and Practice

Although these findings suggest manipulatives are effective at improving mathematical outcomes for students identified or at risk of being identified with a disability, there are limitations that need to be considered when evaluating the results. First, to aggregate ESs, only Phase A (i.e., baseline) and Phase B (i.e., first intervention phase) data were included in the analyses. Thus, if a study evaluated the CRA framework and included Phase B (concrete), Phase C (representation), and Phase D (abstract) data, only the concrete phase data were included. Second, all the alternating treatment designs were comparing concrete with virtual manipulatives, which were target interventions for the current meta-analysis. An ES comparing treatments would not have been representative of effects reported against baseline; thus, these studies were not included in the omnibus ES calculation. Third, the generalizability of this research is limited because only SCEDs were included, resulting in all included participants being identified or at risk of being identified with a disability. Finally, although gray literature was not excluded during screening, the systematic search only identified nine studies (all dissertations) that were not peer-reviewed publications; thus, the omnibus ES is likely inflated because null findings are less likely to be published (Hopewell, Loudon, Clarke, Oxman, & Dickersin, 2009).

In future studies, researchers should seek to consult accepted quality standards from the field (e.g., U.S. Department of Education, Institute of Education Sciences, WWC, 2017; or Council of Exceptional Children [CEC], Cook et al., 2015) when designing research projects. Nearly a third of the data set failed to meet quality indicators, indicating the external validity of findings may be limited. Additional research that aims to systematically replicate existing work to varying participant demographics (i.e., age and disability category) will provide practitioners more informed decisions on intervention selection and implementation. Although, we found age was not a factor associated with the effectiveness of manipulatives, less research exists on manipulatives students use in preschool settings and intermediate settings. Finally, research that aims to examine manipulatives' impact on long-term retention (i.e., maintenance) and generalization is needed, given differential findings reported by Carbonneau et al. (2013) as well as limited current attention to the issue of maintenance and generalization from use of manipulatives for students identified or at risk of being identified with a disability, as compared with research on acquisition.

Results from the current investigation hold implications for practice. First, evidence exists regarding the use of manipulatives for mathematics instruction, suggesting manipulatives will improve the mathematical outcomes for students identified or at risk of being identified with a disability across a variety of mathematical domains. The authors want to emphasize that all but one of the 48 studies (i.e., Scheuermann, Deshler, & Schumaker, 2009) used explicit instruction when incorporating manipulatives; thus, it is recommended to use manipulatives within an explicit instruction framework for this population of students. Second, a variety of concrete manipulatives were used: those purchased for the intended usage (e.g., base 10 blocks) and those made by teachers or repurposed from readily available materials (e.g., paper plates, paper clips), suggesting it may not be important for teachers to have specifically purchased manipulatives to provide the instructional experience for students. In addition, given the positive effects for virtual manipulatives (Satsangi & Miller, 2017), practitioners should consider incorporating virtual manipulatives into mathematics lessons. This may be particularly true for secondary students, as virtual manipulatives can provide less stigmatization than concrete manipulatives as well as be more socially desirable (Bouck, Working, & Bone, 2018).

Conclusion

The purpose of this meta-analysis was to examine the effectiveness of interventions using manipulatives on the mathematical outcomes for students at risk or identified with a disability. Findings suggest the use of manipulatives for mathematics instruction is successful across implementers, participant characteristics, and intervention characteristics. The one caveat is students identified with an emotional or behavioral disorder received less benefits. Future research should examine whether the inclusion of behavioral supports alongside manipulatives yields more favorable outcomes.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education through Grant R324B160038 awarded to UNC-Chapel Hill, by the Office of Special Education Programs. The opinions expressed within this manuscript represent those of the authors and do not represent the U.S. Department of Education.

Supplemental Material

Supplemental material for this article is available online.

References

Arizona Software. (2010). *GraphClick*. Retrieved from http://www.arizona-software.ch/graphclick/

Baron, A., & Derenne, A. (2000). Quantitative summaries of single-subject studies: What do group comparisons tell us about individual performances? *The Behavior Analyst*, *23*, 101–106. doi:10.1007/BF03392004

Becker, B. J. (2005). Failsafe *N* or file-drawer number. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 111–125). Hoboken, NJ: Wiley.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*, 1088–1101.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis 2.0*. Englewood, NJ: Biostat.

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

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*, 65–106. doi:10.1353/etc.2018.0003

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

Bouck, E. C., Working, C., & Bone, E. (2018). Manipulative apps to support students with disabilities in mathematics. *Intervention in School and Clinic*, *53*, 177–182. doi:10.117/1053451217702115

Boyle, M. A., Samaha, A. L., Rodewald, A. M., & Hoffmann, A. N. (2013). Evaluation of the reliability and validity of GraphClick as a data extraction program. *Computers in Human Behavior*, *29*, 1023–1027. doi:10.1016.j.chb.2012.07.031

Brossart, D. F., Vannest, K. J., Davis, J. L., & Patience, M. A. (2014). Incorporating nonoverlap indices with visual analysis for quantifying intervention effectiveness in single-case experimental designs. *Neuropsychological Rehabilitation*, *24*, 464–491. doi:10.1080/09602011.2013.868361

Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist*, *19*, 1–15. doi:10.1037/h0044160

Burns, M. K. (2012). Meta-analysis of single-case design research: Introduction to the special issue. *Journal of Behavioral Education*, *21*, 175–184. doi:10.1007/s10864-012-9158-9

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. Kratochwill & J. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 187–212). New York, NY: Routledge.

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*, 380–400. doi:10.1037/a0031084

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, *10*, 101–129.

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

Cooper, H. (2016). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). Thousand Oaks, CA: SAGE.

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:10.1177/001440230507100202

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*, 224–239. doi:10.1002/jrsm.1052

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods*, *4*, 324–341. doi:10.1002/jrsm.1086

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558. doi:10.1002/sim.1186

Hopewell, S., Loudon, K., Clarke, M. J., Oxman, A. D., & Dickersin, K. (2009). Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database of Systematic Reviews*, *90*, 1631–1640.

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

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. F. (2009). Concrete instantiations of mathematics: A double-edged sword. *Journal for Research in Mathematics Education*, *40*, 90–93.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *What Works Clearinghouse: Single-case design technical documentation*. Retrieved from https://files.eric.ed.gov/fulltext/ED510743.pdf

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

Losinski, M. L., Ennis, R. P., Sanders, S. A., & Nelson, J. A. (2019). A meta-analysis examining the evidence-base of mathematical interventions for students with emotional disturbances. *The Journal of Special Education*, *52*, 228–241. doi:10.1177/0022466918796200

Maggin, D. M., Pustejovsky, J. E., & Johnson, A. H. (2017). A meta-analysis of school-based group contingency interventions for students with challenging behavior: An update. *Remedial and Special Education*, *38*, 353–370. doi:10.1177/0741932517716900

Manolov, R., Guilera, G., & Solanas, A. (2017). Issues and advances in the systematic review of single-case research: A commentary on the exemplars. *Remedial and Special Education*, *38*, 387–393. doi:10.1177/0741932517726143

McNeil, N. M., Uttal, D. H., Jarvin, L., & Sternberg, R. J. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning and Instruction*, *19*, 171–184. doi:10.1016/j.learninstruc.2008.03.005

Morin, K. L., Ganz, J. B., Gregori, E. V., Foster, M. J., Gerow, S. L., Genc-Tosun, D., & Hong, E. R. (2018). A systematic quality review of high-tech AAC interventions as an evidence-based practice. *Augmentative and Alternative Communication*, *34*, 104–117. doi:10.1080/07434618.2018.1458900

Moyer, P. S. (2001). Are we having fun yet? How teachers use manipulatives to teach mathematics. *Educational Studies in Mathematics*, *47*, 175–197.

National Assessment of Educational Progress. (2017). *The nation's report card 2017: Mathematics and reading assessments*. Available from https://www.nationsreportcard.gov/

National Mathematics Advisory Panel. (2007). *Preliminary report: National mathematics advisory panel*. Retrieved from https://ed.gov/about/bdscomm/list/mathpanel/pre-report.pdf

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, *35*, 303–322. doi:10.1177/0145445511399147

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*, 284–299. doi:10.1016/j.beth.2010.08.006

Piaget, J. (1962). *Play, dreams, and imitation in childhood*. New York, NY: W.W. Norton.

Pustejovsky, J. E. (2016). *Between-case standardized mean difference estimator*. Retrieved from https://jepusto.shinyapps.io/scdhlm/

Pustejovsky, J. E. (2018). Procedural sensitivities of effect sizes for single-case designs with directly observed behavioral outcome measures. *Psychological Methods, 24*(2), 217-235. Advance online publication. doi:10.1037/met0000179

Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics*, *39*, 368–393. doi:10.3102/1076998614547577

Riley-Tillman, T. C., & Burns, M. K. (2009). *Single case design for measuring response to educational intervention*. New York, NY: Guilford.

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

Rosnow, R. L., & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276–1284.

Salzberg, C. L., Strain, P. S., & Baer, D. M. (1987). Meta-analysis for single-subject research: When does it clarify, when does it obscure? *Remedial and Special Education*, *8*, 43–48. doi:10.1177/074193258700800209

Sarama, J., & Clements, D. H. (2016). Physical and virtual manipulatives: What is "concrete?" In P. S. Moyer-Packenham (Ed.), *International perspectives on teaching and learning mathematics with virtual manipulatives* (pp. 71–93). Basel, Switzerland: Springer International.

Satsangi, R., & Miller, B. (2017). The case for adopting virtual manipulatives in mathematics education for students with disabilities. *Preventing School Failure: Alternative Education for Children and Youth*, *61*, 303–310. doi:10.1080/1045988X.2016.1275505

Scheuermann, A. M., Deshlet, D. D., & Schumaker, J. B. (2009). The effects of explicit inquiry routines on the performance of students with learning disabilities on one-variable equations. *Learning Disabilities Quarterly*, *32*(2), 103-120.

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

Test, D. W., Mazzotti, V. L., Fowler, C. H., Kortering, L., & Kohler, P. (2009). Evidence-based secondary transition predictors for improving postschool outcomes for students with disabilities. *Career Development for Exceptional Individuals*, *32*(3), 160-181.

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2017). *What Works Clearinghouse: Procedures handbook version 4.0*. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf

Vannest, K. J., Parker, R. I., Gonen, O., & Adiguzel, T. (2016). *Single case research: Web based calculators for SCR analysis*. Available from http://www.singlecaseresearch.org/

Vannest, K. J., Peltier, C., & Haas, A. (2018). Results reporting in single case experiments and single case meta-analysis. *Research in Developmental Disabilities*, *79*, 10–18. doi:10.1016/j.ridd.2018.04.029