Journal of Educational Psychology
2013, Vol. 105, No. 2, 380–400

0022-0663/13/$12.00 DOI: 10.1037/a0031084

A Meta-Analysis of the Efficacy of Teaching Mathematics With Concrete Manipulatives

Kira J. Carbonneau, Scott C. Marley, and James P. Selig
University of New Mexico

The use of manipulatives to teach mathematics is often prescribed as an efficacious teaching strategy. To examine the empirical evidence regarding the use of manipulatives during mathematics instruction, we conducted a systematic search of the literature. This search identified 55 studies that compared instruction with manipulatives to a control condition where math instruction was provided with only abstract math symbols. The sample of studies included students from kindergarten to college level ($N = 7{,}237$). Statistically significant results were identified with small to moderate effect sizes, as measured by Cohen's *d*, in favor of the use of manipulatives when compared with instruction that only used abstract math symbols. However, the relationship between teaching mathematics with concrete manipulatives and student learning was moderated by both instructional and methodological characteristics of the studies. Additionally, separate analyses conducted for specific learning outcomes of retention ($k = 53$, $N = 7{,}140$), problem solving ($k = 9$, $N = 477$), transfer ($k = 13$, $N = 3{,}453$), and justification ($k = 2$, $N = 109$) revealed moderate to large effects on retention and small effects on problem solving, transfer, and justification in favor of using manipulatives over abstract math symbols.

Keywords: mathematics, manipulatives, concrete objects, activity-based learning, hands-on learning

Results from the 2011 National Assessment of Educational Progress (National Center for Education Statistics, 2011) indicate 60% of fourth-grade and 57% of eighth-grade United States students failed to meet standards of proficiency in mathematics. Furthermore, with only 10% of fourth graders and 6% of eighth graders meeting international standards of advanced proficiency, U.S. students rank below their same-age peers from eight countries (National Center for Education Statistics, 2008). These results, and comparable findings from prior years, have provided President Obama motivation for a recent executive branch initiative known as Educate to Innovate (The White House, Office of Press Secretary, 2009). This initiative was developed to target student achievement within science, technology, engineering, and math education with a focus on increasing domain-specific critical reasoning skills. If the goal of Educate to Innovate is to help students reach high levels of mathematics achievement, efficacious instructional strategies need to be identified. Therefore, a careful examination of contemporary instructional strategies is necessary to identify strategies that improve mathematics achievement.

Instructional strategies that use manipulatives are often suggested as effective approaches to improve student mathematics achievement (Gürbüz, 2010; Sherman & Bisanz, 2009). Math manipulative-based instructional techniques are approaches that include opportunities for students to physically interact with objects to learn target information (Carbonneau & Marley, 2012). As examples, at the elementary level, teachers use play money to help students learn basic arithmetic functions, and at the high school level, teachers use plastic algebra tiles to teach concepts associated with division and multiplication within an equation. The National Council of Teachers of Mathematics (NCTM, 2000) has recommended that students be provided access to manipulatives in order to develop mathematical understanding. In addition, teacher education textbooks often contain sections suggesting that teachers use manipulatives during mathematics instruction (e.g., Billstein, Libeskind, & Lott, 2009; Copley, 2000). In the cases of national organizations and textbooks, when an instructional strategy is prescribed to a professional audience, an underlying assumption is that sound scientific evidence supports the recommendation. However, evidence supporting the efficacy of concrete math manipulatives is inconsistent. Specifically, the efficacy of manipulatives in mathematics instruction has not been uniformly observed with various populations, math topics, and cognitive outcomes. Variability of the effectiveness within instructional strategies of this nature may result in the misapplication of the instructional technique.

The instructional strategies literature is not definitive regarding the efficacy of concrete manipulatives. Studies have found that using manipulatives in math instruction, when compared with instruction that did not use manipulatives, may benefit student

This article was published Online First December 17, 2012.

Kira J. Carbonneau, Scott C. Marley, and James P. Selig, Department of Individual, Family and Community Education, Educational Psychology Program, College of Education, University of New Mexico.

Funding for this research was provided by the University of New Mexico's College of Education. Preliminary results were presented at the American Educational Research Association 2012 Annual Conference, Vancouver, Canada. We thank Carolyn J. Hushman for her assistance with coding studies.

Correspondence concerning this article should be addressed to Kira J. Carbonneau or Scott C. Marley, Department of Individual, Family and Community Education, Educational Psychology Program, College of Education, University of New Mexico, 117 Simpson Hall, MSC 05 3040, Albuquerque, NM 87131-1246. E-mail: kjcarbonneau@gmail.com or marley@unm.edu

learning (Gürbüz, 2010), result in comparable performance (Canobi, 2005; Dyer, 1996), or reduce student learning (Shoecraft, 1971). These contradictions may exist as a result of systematic factors. For instance, the level of instructional guidance, type of manipulative, age of learners, and other characteristics of a learning environment may impact the effectiveness of the intervention. Therefore, a systematic review of the math manipulatives literature is necessary to understand the variations in results observed between studies.

Sowell (1989) performed the first research synthesis of the manipulatives literature with a meta-analysis of the use of manipulatives applied to mathematics learning. Sowell's results suggested that relative to studies that did not use manipulatives, small-sized statistical differences in favor of the use of manipulatives on measures of recall were present when instruction was implemented over a school year. One limitation of Sowell's synthesis is that it did not examine whether instructional characteristics moderate the effectiveness of math manipulatives in terms of student learning. In addition to this limitation, there has been a considerable expansion of the math manipulatives literature base since Sowell's 1989 study. Both of these circumstances justify another systematic review of the math manipulatives literature. Therefore, the purposes of the present meta-analysis were to determine the average effect of math manipulatives and to identify potential moderators of the effectiveness of manipulatives. Instructional moderators were identified based on theoretical explanations for the efficacy of manipulatives, whereas methodological characteristics were examined to evaluate the credibility of the literature (Marley & Levin, 2011).

Moderators of the Efficacy of Manipulatives: Instructional Characteristics

Potential instructional moderators of the efficacy of teaching with manipulatives can be derived from contemporary human development and cognitive theories (McNeil & Jarvin, 2007). According to these theoretical explanations, concrete manipulatives facilitate learning by (a) supporting the development of abstract reasoning (Bruner, 1964; Montessori, 1964; Piaget, 1962), (b) stimulating learners' real-world knowledge (Baranes, Perry, & Stigler, 1989; Rittle-Johnson & Koedinger, 2005), (c) providing the learner with an opportunity to enact the concept for improved encoding (Kormi-Nouri, Nyberg, & Nilsson, 1994), and (d) affording opportunities for learners to discover mathematical concepts through learner-driven exploration (Bruner, 1961; Papert, 1980; Piaget & Coltman, 1974). Each of these theoretical explanations provides instructional characteristics that may reduce or increase the effectiveness of math manipulatives. The following sections describe each theoretical explanation and theory-relevant instructional factors.

Development of Abstract Reasoning

According to developmental theorists (Bruner, 1964; Montessori, 1964; Piaget, 1962), young children are expected to obtain cognitive benefits from exploring mathematical concepts with manipulatives. Empirical research examining concrete manipulatives are commonly situated through Piagetian developmental statuses with a focus on children at ages associated with concrete operations (Fennema, 1972; Fujimura, 2001; Garcia, 2004). In addition, theoretical perspectives indicate that children in early childhood (age 7 and younger) should benefit from exploring mathematical concepts with manipulatives (Montessori, 1964). Children within both age groups are expected to derive greater cognitive benefits from manipulatives relative to older children (Fennema, 1972; Resnick & Omanson, 1987). The reason for this expectation is that younger children are assumed to have a greater dependency on physically interacting with their environment to construct meaning (Bruner, 1964; Piaget & Coltman, 1974). Through these physical interactions with the environment, young children are expected to gain proficiency with higher level representations in a predictable sequence. This sequence predicts that the ability for children to capitalize from visual representations should precede symbolic representations.

Inherent in these theoretical perspectives is the prediction that the developmental status of students should moderate the efficacy of teaching math topics with concrete manipulatives. It is expected that older students who have developed the ability to reason abstractly can benefit from instruction that consists exclusively of symbolic representations. Younger children, however, are predicted to experience more difficulty when provided instruction that solely consists of symbolic representation. Therefore, the assumed cognitive benefits of manipulating concrete objects to represent mathematical concepts should be greater for younger children who are still developing proficiency with higher level representations.

Stimulating Real-World Knowledge

The use of manipulatives in mathematics instruction has been cited as a strategy to allow students to draw on their practical knowledge (Burns, 1996). This line of reasoning suggests that concrete objects that resemble everyday items should assist students in making connections between abstract mathematical concepts and the real world (Brown, McNeil, & Glenberg, 2009). Support for this argument is provided by evidence indicating that when prior knowledge of a concept is partial, or absent, providing known concrete objects may help learners construct context-relevant schemas (e.g., Tindall-Ford & Sweller, 2006). However, results from empirical research examining the connection between student learning and the type of manipulatives used during instruction has been counterintuitive and could potentially account for some of the inconsistencies within the manipulation-based literature. Research examining different types of manipulatives tends to focus on the perceptual richness of concrete objects and how details of an object may hinder or aid learning.

Research examining the perceptual richness of a manipulative has primarily focused on realism or visual details of manipulatives. These examinations compare realistic manipulatives (e.g., manipulatives that look like pizza or money) that are perceptually rich (McNeil, Uttal, Jarvin, & Sternberg, 2009) to manipulatives that are nondescript or bland in nature (e.g., manipulatives that represent geometric shapes or place value). Results from these studies suggest that perceptually rich manipulatives may hinder learning of targeted mathematics concepts and/or performance solving mathematics problems (Kaminski, Sloutsky, & Heckler, 2009; McNeil et al., 2009). Explanations for why learning is inhibited by perceptually rich manipulatives have focused on generalizing learning to other contexts (Martin, 2009), surface information that is irrelevant to the target concept (Kaminski et al., 2009; McNeil et al., 2009), and children not recognizing that concrete objects can be representative of an actual object and abstract mathematics concepts (Uttal, O'Doherty, Newland, Hand, & DeLoache, 2009).

Enactment Effects

Instructional strategies that use manipulatives may be effective because of physical enactment. In other words, the encoding and subsequent retrieval of target information may occur via nonverbal coding or a motoric channel. A well-developed literature exists examining what are known as self-performed tasks (SPTs). SPTs are tasks that participants physically perform during a learning activity. Often examined in paired-associate or list-learning contexts, SPTs have been found to result in robust encoding that enhances subsequent retrieval of target information (Engelkamp, Zimmer, Mohr, & Sellen, 1994; Kormi-Nouri, Nyberg, & Nilsson, 1994).

Dual coding theory offers an explanation for the memory benefits of SPTs. According to this theory, verbal and nonverbal representations are stored in separate but connected stores in long-term memory (Paivio, 1986). Consequently, it is proposed that activation of one form of representation leads to the activation of the other, resulting in the improved retrieval of target information (Clark & Paivio, 1987). In an instructional context, a child studying math facts using manipulatives to represent quantities is learning the target concept with both forms of representation present. Later, when asked to remember the target information, the child would have access to a verbal code consisting of target math facts and a nonverbal code consisting of interactions with the manipulatives. The successful retrieval of one form of representation is expected to activate the other, which in turn should result in greater performance on learning outcomes (for relevant discussion, see Marley & Levin, 2006).

The process of enactment has been demonstrated as an efficacious learning strategy within the content area of reading comprehension. Manipulating objects as directed by a narrative has been found to improve memory for spatial relationships and story events (Biazak, Marley, & Levin, 2010). Likewise, the ability to benefit from imagery instruction has been associated with the enactment of manipulatives. For example, Glenberg, Gutierrez, Levin, Japuntich, and Kaschak (2004) found that interacting with text-relevant objects during reading instruction increased comprehension of stories when participants were subsequently asked to imagine manipulating the text-relevant objects. This finding has been replicated in studies with children from samples representing diverse populations (Marley, Szabo, Levin, & Glenberg, 2011; Marley & Szabo, 2010).

A potential problem with physical enactment has been identified by several authors (Martin, 2009; Sarama & Clements, 2009). The simple act of moving manipulatives is likely not sufficient for promoting learning. Without explicit instruction, children may not move objects in a manner that appropriately represents the mathematics concept being taught. In other words, the instructional guidance provided is expected to influence the efficacy of manipulatives and the process of engaging in SPTs.

Learner-Driven Exploration

Many have suggested that providing learners with opportunities to discover mathematical concepts through unstructured learner-driven exploration will result in robust learning outcomes (Bruner, 1961; Piaget & Coltman, 1974). These theorists propose that learners are better able to construct meaningful knowledge when given opportunities to discover concepts (Lefrançois, 1997). Empirical research has provided evidence contradicting this notion with results indicating that providing learners with instructional guidance on topics rather than allowing them to work within a purely unstructured context results in higher levels of student performance (Mayer, 2004).

Instructional guidance offered to learners can be defined as the amount of instructional support provided during the learning process and falls on a continuum of student- versus teacher-controlled learning (Kirschner, Sweller, & Clark, 2006; Mayer, 2004). On one end of this continuum are student-controlled strategies that allow learners to use manipulatives in an unstructured or less structured environment (i.e., a low guidance or discovery learning environment). Students in low guidance scenarios, often identified as math explorations, are given manipulatives with little or no instruction on how to manipulate the objects to represent mathematical concepts under study (Hinzman, 1997; Kuhfittig, 1974; LeBlanc, 1968). On the other end of the continuum are teacher-controlled strategies in which students interact with manipulatives as instructed by a teacher (i.e., direct instruction).

A recent synthesis of the instructional guidance literature indicates the provision of instructional guidance results in greater performance on learning outcomes relative to pure discovery (Alfieri, Brooks, Aldrich, & Tenenbaum, 2011). Reading and listening strategy research further supports the importance of instructional guidance when using concrete manipulatives (Glenberg, Brown, & Levin, 2007; Marley, Levin, & Glenberg, 2007, 2010; Marley et al., 2011). However, Martin (2009) warned that too much instructional guidance with concrete manipulatives can impede learning by confining students to interpretations that do not transfer to novel circumstances. If this so, it is expected that the provision of high instructional guidance with manipulatives will result in lower performance on outcomes related to transfer of learning.

Moderators of the Efficacy of Manipulatives: Methodological Characteristics

Methodological aspects of a study affect the credibility of the claims that can be made regarding the causal relationship between an instructional strategy and beneficial learning outcomes (for relevant discussions, see Marley & Levin, 2011; Shadish, Cook, & Campbell, 2002). The robustness of claims derived from a literature can be assessed by examining the prevalence of studies that have characteristics linked with limitations in internal (e.g., pre- and postdesigns) and external (e.g., researcher-implemented treatments) validities. Whether these factors are associated with effect size is of interest, because educational recommendations should be based upon a body of scientifically credible evidence. For the present meta-analytic review, potential moderators of interest were whether design (pre- and postdesign, quasi-experiment, true experiment), type of test (standardized test, researcher-created tests), assumption of statistical independence (accounted for in analysis, unaccounted for in analysis), implementer (researcher, teacher), and peer review status (published, unpublished) were associated with effect size. Although peer review status is not a methodological characteristic, it is included because acceptance in a peer-reviewed publication may serve as a proxy for a study's methodological rigor.

Present Study

The purpose of the present study was to ascertain the effectiveness of using manipulatives to teach mathematics when compared with teaching mathematics with only abstract math symbols. Moderators were examined to determine whether efficacy of this strategy differed by instructional and methodological characteristics. In summary, we sought to answer the following research questions:

1. What is the average effect of using concrete manipulatives in mathematics instruction?

2. Does the relationship between using concrete manipulatives and student learning vary by learning outcome?

3. Do instructional characteristics of studies moderate the relationship between using concrete manipulatives and learning outcomes?

4. Do methodological characteristics of studies moderate the relationship between using concrete manipulatives and learning outcomes?

In addition to addressing the primary research questions, we evaluated the overall quality of the literature by describing the prevalence of specific design characteristics. For a literature base to be considered robust in terms of validity of causal inferences and generalizability, it should primarily consist of studies that begin with random assignment and are representative of instructional contexts.

Method

Literature Search

An exhaustive search for studies on manipulatives and mathematics was performed between August 2010 and March 2011 with relevant keywords and their combinations (e.g., *mathematics*, *manipulatives*, *concrete objects*, *activity-based learning*, *hands-on learning*) with six major databases in the social sciences (Education Resources Information Center, Education Research Complete, PsycARTICLES, PsycINFO, JSTOR, and ProQuest Digital Dissertations). This search resulted in the identification of 94 articles. From these articles, ancestral and secondary citations from the studies' references were collected and examined for relevance to the present study. The search of ancestral and secondary citations resulted in 102 additional references ($k = 196$). Of the 196 studies, projects that did not report empirical findings were removed. This resulted in 101 studies to be reviewed for inclusion in the meta-analysis.

Criteria for Inclusion

Four conditions were established to restrict studies to those that empirically examined the efficacy of manipulatives in mathematics instruction. First, to be included, a study was required to compare an instructional technique that used manipulatives with a comparison group that taught math with only abstract math symbols. This comparison group was defined by the following attributes: (a) no manipulatives were present, (b) all students were taught the same math concept, and (c) no iconic representations (e.g., pictures of base-10 blocks or virtual manipulatives) were present. Of the 101 studies, 21 (20.7%) failed to meet these criteria. This inclusion criterion produces a very specific comparison in which conditions where students physically interacted with concrete objects were compared with conditions where students were solely taught mathematics concepts with abstract math symbols.

Second, to be included in the meta-analysis, the examined instructional treatments must have provided some form of instruction during which students were able to learn from the manipulatives. Studies that examined only the performance of students with manipulatives were excluded; four studies (3.9%) failed to meet this criterion. The third criterion is based on the definition of manipulatives; studies that required students to work with rulers, scales, or calculators were not included, as these were seen as tools rather than manipulatives (8.0%, $k = 9$). Lastly, studies had to provide sufficient quantitative information to estimate an effect size, which resulted in the elimination of 11 studies (10.8%). Additionally, two publications reported results from the same study; therefore only one was retained in the sample. The screening process resulted in a total of 55 studies upon which all meta-analytic procedures were conducted.[1] Studies meeting the inclusion criteria along with summative information on design and findings are presented in Table 1.

Study Coding

All 55 studies were coded with a standardized protocol. The protocol was developed iteratively as studies were accumulated to include the moderators of interest. After a final coding scheme was developed, studies were coded for the characteristics described below. Two raters independently coded an overlapping random sample of 18 studies (32%) to assure consistency in coding. Interrater agreement, as measured by Cohen's κ for the categorical variables, ranged from .82 to 1.0. Pearson's r for continuous variables ranged from .89 to 1.0.

Effect size. Cohen's *d* (1988), a measure of effect size, was calculated for each study in the meta-analysis. Cohen's *d* values are obtained by dividing the difference of the treatment means by the pooled standard deviation. This measure of effect size is commonly reported in studies examining the effect of a manipulated independent variable (e.g., manipulatives vs. control) on a continuous dependent variable (e.g., retention, problem solving, transfer, or justification). When study statistics were not directly

[1] Sowell's 1989 meta-analysis of the math manipulative research consisted of 60 studies, of which approximately 23 contrasted pictures versus a symbolic representation. The publication does not include a list of reviewed studies. Sowell kindly responded to a request for a list of reviewed studies. She no longer had access to the list.

Table 1
Summary of Study Characteristics

Study	N	Days	Design[a]	Coded instructional characteristics[b]	Mean Cohen's d	Main findings
Aburime (2007)	185	50	QEX	FO, PRM, HG, GE, GI	0.01	Students in high school taught geometry with manipulatives performed the same on a measure of retention as those not taught with manipulatives.
Anderson (1957)	541	40	QEX	FO, LG, AR, GI,	0.08	Eighth-grade students taught algebra with manipulatives performed the same on a measure of retention as those taught with a textbook.
Aurich (1963)	90	180	QEX	PO, BM, HG, AR, GI	0.89	First graders taught arithmetic with Cuisenaire rods performed better on measures of reasoning and retention than those taught with a textbook.
Babb (1975)	76	25	QEX	CO, PRM, LG, AR, GI	−0.05	Second graders taught arithmetic with manipulatives performed the same on a measure of retention as those taught without manipulatives.
Battle (2007)	16	5	QEX	PO, PRM, LG, AR, GI	−0.93	First graders taught addition and subtraction with counters performed worse on a measure of retention than those taught without counters.
Bring (1972)	102	15	EX	CO, PRM, LG, AL, II	0.28	Fifth and sixth graders taught algebra with manipulatives performed better on a measure of retention than those students taught with a textbook.
Butler et al. (2003)	50	10	WS	CO, BM, HG, FR, GI	2.24	Mathematic disabled seventh graders taught fractions with manipulatives answered more items correctly on a postmeasure of retention.
Carmody (1970)	96	11	QEX	CO, PRM, HG, AR, GI	0.43	Sixth graders taught arithmetic with manipulatives performed better on a measure of transfer than students taught without manipulatives. Performance on a measure of retention was the same for both groups.
Cook (1967)	66	120	QEX	PO, LG, AR, GI	0.08	First-grade students taught arithmetic with manipulatives performed the same on a measure of problem solving than those taught with a textbook.
Cramer et al. (2002)	1,666	30	QEX	CO, PRM, HG, FR, GI	0.88	Fourth- and fifth-grade students taught fractions with manipulatives performed better on a measure of retention than those taught with a textbook.
Dawson (1955)	280	22	QEX	CO, PRM, LG, AR, GI	0.58	Fourth-grade students taught division with manipulatives performed better on a measure of retention than students who were taught without manipulatives.
Dyer (1996)	90	Unknown	QEX	FO, BM, LG, AL, GI	0.00	College-level students taught algebra with algebra tiles performed the same on a measure of retention than students taught with a textbook.
Egan (1990)	81	180	QEX	CO, BM, LG, AR, GI	−0.30	Second graders taught arithmetic with Cuisenaire rods performed worse on a measure of retention than those who were taught with a textbook.
Ekman (1967)	196	18	QEX	CO, PRM, HG, AR, GI	0.09	Third graders taught arithmetic with manipulatives performed the same on measures of retention and transfer as those taught with a textbook.

Table 1 (*continued*)

Study	*N*	Days	Design[a]	Coded instructional characteristics[b]	Mean Cohen's *d*	Main findings
Fennema (1972)	95	14	EX	CO, BM, HG, AR, GI	−0.65	Second graders taught arithmetic with manipulatives performed worse on a measure of transfer than students taught with a textbook.
Fujimura (2001)	76	1	EX	CO, PRM, LG, AL, II	0.73	Fourth graders taught proportions with manipulatives performed better on a measure of retention than those taught without manipulatives.
Garcia (2004)	64	20	QEX	CO, LG, AR, GI	−0.14	Third and fourth graders taught arithmetic with manipulatives performed the same on a measure of retention as those taught without manipulatives.
Getgood (2000)	287	10	QEX	CO, BM, HG, AL, GI	0.07	Sixth graders taught algebra with factor blocks performed better on a measure of immediate retention than those taught with a textbook. Performance on a second measure of retention was the same for both groups.
Goins (2001)	30	Unknown	QEX	FO, BM, HG, AL, GI	1.21	Ninth-grade students who were taught algebra with algebra tiles performed better on a postassessment than students who did not have access to the tiles.
Gürbüz (2010)	80	7	EX	CO, BM, HG, FR, GI	3.11	Seventh-grade students taught fractions with manipulatives performed better at word problems pertaining to probability than students taught without manipulatives.
Hawkins (1982)	35	20	WS	CO, PRM, HG, FR, GI	1.60	Third-grade students taught fractions with manipulatives answered more items correctly on a postmeasure of retention.
Hiebert et al. (1991)	25	11	WS	CO, BM, HG, PV, GI	0.47	Fourth-grade students taught decimals with manipulatives answered more items correctly on a postmeasure of retention with manipulatives. This gain was not seen on a measure of retention performed without manipulatives.
Hinzman (1997)	34	45	QEX	FO, BM, LG, AL, GI	0.20	College-level students taught algebra with manipulatives performed the same on a measure of retention than students who were not provided manipulatives.
Johnson (1970)	64	20	EX	CO, PRM, LG, AL, II	0.99	Fourth- and fifth-grade students taught prealgebra with manipulatives performed better on measures of retention than students taught without manipulatives.
Jordan et al. (1999)	125	Unknown	QEX	CO, PRM, HG, FR, GI	0.78	Fourth-grade students taught fractions with manipulatives performed better on measures of retentions than students who were taught without manipulatives.
King (1976)	134	5.5	EX	CO, PRM, HG, FR, GI	−0.02	Fourth-grade students taught fractions with manipulatives performed the same on measures of retention and transfer as those students taught with a textbook.
Kuhfittig (1974)	40	2	EX	CO, LG, AR, GI	−0.02	Seventh-grade students taught arithmetic with manipulatives performed the same on measures of retention and transfer as those students who were taught without manipulatives.

(*table continues*)

Table 1 (*continued*)

Study	*N*	Days	Design[a]	Coded instructional characteristics[b]	Mean Cohen's *d*	Main findings
Lucas (1966)	104	50	QEX	PO, PRM, LG, AR, GI	0.24	First-grade students taught with attribute blocks performed better on a measure of problem solving than students who were not taught with manipulatives.
Lucow (1964)	254	30	QEX	CO, BM, HG, AR, GI	0.76	Third-grade students taught arithmetic with manipulatives scored higher on a measure of retention than students taught with a textbook.
McClung (1998)	47	45	QEX	FO, BM, LG, AL, GI	−0.70	Tenth- and 11th-grade students taught algebra with manipulatives performed worse on a measure of retention than students taught without manipulatives.
Miller (1964)	114	9	WS	CO, PRM, HG, FR, GI	3.90	Sixth-grade students taught fractions with manipulatives answered more items correctly on a postmeasure of retention.
Moody et al. (1971)	90	20	EX	CO, PRM, HG, AR, GI	−0.07	Third-grade students taught multiplication with manipulatives performed worse on measures of retention and transfer than students taught with a textbook.
Nasca (1966)	45	180	QEX	CO, BM, HG, AR, GI	0.43	Second-grade students taught arithmetic with rods performed the same on measures of retention and transfer as those students taught with a textbook.
Nichols (1972)	267	Unknown	WS	CO, PRM, LG, AR, GI	2.12	Third-grade students taught multiplication with manipulatives answered more items correctly on a postmeasure of retention.
Nickel (1971)	90	30	QEX	CO, PRM, HG, AR, GI	0.40	Fourth-grade students taught arithmetic with manipulatives performed the same as students taught without manipulatives.
Nishida (2007)	78	1	EX	PO, BM, HG, FR, GI	0.11	First-grade students taught fractions with manipulatives performed the same on a measure of retention as students taught without manipulatives.
Norman (1955)	24	10	QEX	CO, HG, AR, GI	1.48	Third-grade students taught division with manipulatives performed better than students taught with a textbook.
Olkun (2003)	93	2	QEX	CO, PRM, LG, GE, II	0.37	Fourth- and fifth-grade students taught geometry with manipulatives preformed better on a measure of retention than students taught without manipulatives.
Paolini (1977)	26	15	WS	PO, LG, AR, GI	0.88	Kindergarten students taught arithmetic with manipulatives answered more items correctly on a postmeasure of retention.
Peterson et al. (1988)	24	9	EX	CO, PRM, HG, PV, GI	0.67	Learning disabled fourth graders taught place value with manipulatives performed better on measures of retention and transfer than students taught without manipulatives.
Prigge (1978)	146	10	QEX	CO, PRM, LG, GE, II	0.59	Third-grade students taught geometry with manipulatives performed better on measures of retention than students taught with a textbook.
Robinson (1978)	119	5	EX	CO, BM, HG, FR, GI	0.27	Fourth graders taught fractions taught fractions with Cuisenaire rods performed the same on a measure of retention as students taught without rods.

Table 1 (*continued*)

Study	*N*	Days	Design[a]	Coded instructional characteristics[b]	Mean Cohen's *d*	Main findings
Shoecraft (1971)	1096	10	EX	FO, LG, FR, GI	−0.04	Seventh- and ninth-grade students taught fractions with manipulatives performed the same on measures of retention and problem solving as students taught without manipulatives.
Slaughter (1980)	217	Unknown	QEX	CO, BM, LG, AR, GI	0.18	Third- and fifth-grade taught arithmetic with manipulatives performed the same on measures of retention as students taught with a textbook.
Smith & Montani (2008)	12	12	WS	CO, BM, HG, AR, GI	0.84	Third graders taught multiplication with manipulatives answered more questions correctly on a postmeasure or retention.
Threadgill-Sowder & Juilfs (1980)	147	3	EX	CO, PRM, LG, GE, GI	0.47	Seventh-grade students taught geometry with manipulatives performed better on a measure of retention than students taught without. Performance on a measure of transfer favored students who were taught without manipulatives.
Steen et al. (2006)	31	15	WS	PO, HG, GE, GI	1.70	First-grade students taught geometry with manipulatives answered more items correctly on a postmeasure of retention.
Steger (1977)	52	12	WS	PO, BM, HG, PV, GI	0.63	First-grade students taught place value with manipulatives answered more items correctly on a postmeasure of retention.
Suh & Moyer (2007)	36	5	WS	CO, MB, HG, FR, GI	2.76	Third-grade students taught fractions with manipulatives answered more items correctly on a postmeasure of retention.
Taylor (2001)	58	3	QEX	CO, PRM, LG, AL, GI	−0.96	Fifth-grade students taught probability with manipulatives performed worse on a measure of retention than students taught with a textbook. Performance on a transfer measure was the same for both groups.
Wallace (1974)	154	Unknown	QEX	CO, BM, HG, AR, GI	1.42	Fifth-grade students taught arithmetic with Cuisenaire rods performed better on a measure of retention than students taught with a textbook.
Weber (1970)	30	30	QEX	PO, PRM, HG, AR, GI	0.18	First-grade students taught arithmetic with manipulatives performed the same on a measure of retention as students taught with a textbook.
Witzel et al. (2003)	68	20	QEX	CO, BM, HG, FR, GI	0.68	Sixth- and seventh-grade students taught fractions with manipulatives performed better on a measure of retention than students taught without manipulatives.
Wood (1974)	40	4	WS	CO, PRM, LG, AR, GI	0.12	Second-grade students taught multiplication with manipulatives answered the same amount of questions correctly on a postmeasure of retention and transfer.
Yuan et al. (2010)	60	4	WS	FO, PRM, LG, GE, GI	0.72	Eighth-grade students taught geometry with manipulatives answered more questions correctly on a postassessment of problem solving.

[a] EX = experiment; QEX = quasi-experiment; WS = within subjects; [b] PO = preoperational students; CO = concrete operational students; FO = formal operational students; PRM = perceptually rich manipulatives; BM = bland manipulatives; HG = high instructional guidance; LG = low instructional guidance; PV = place value; AR = arithmetic; GE = geometry; FR = fractions; AL = algebra; GI = group instruction; II = individual instruction.

reported, Cohen's *d* values were calculated with reported descriptive statistics or observed *F* or *t* statistics (Rosenthal, 1984). Given that studies with larger samples should have a more precise estimate of the effect of manipulatives, studies were weighted to allow large samples to have more influence. Weights for each study were calculated from the reciprocal of the computed variance for *d* (for details, see Lipsey & Wilson, 2001).

A number of studies measured several outcomes of interest. Studies with various outcomes allowed for the coding of multiple effect sizes. For example, the study by Aurich (1963) measured three of the four coded learning outcomes (retention, problem solving, and transfer). This afforded the calculation of three effect sizes. To avoid the potential for nonindependence of effect sizes, effect sizes from the same study were averaged (Rosenthal & Rubin, 1986) to extract one effect size for each study. Effect sizes were also disaggregated to examine differential effects of manipulatives across learning outcomes.

Instructional Moderators

Developmental status. Both age and grade level were coded from each study. In cases with more than one grade level or age present, the mean was recorded. Subsequently, the variable of age was grouped into three categories equivalent to Piaget's stages of development, with samples of students from ages 3–6 coded as preoperational, ages 7–11 coded as concrete operational, and age 12 and older coded as formal operational.

Perceptual richness. Manipulatives were coded as either perceptually rich or bland in nature. A perceptually rich manipulative is defined as an object that is either representative of a real object or the actual object. For example, toy pizzas were coded as a perceptually rich manipulatives (e.g., Ekman, 1967; Peterson, Mercer, & O'Shea, 1988). Bland manipulatives included objects that are nondescript such as plain rectangular blocks or tangrams (e.g., Dyer, 1996; Egan, 1990).

Level of instructional guidance. Support provided to students was coded to represent two levels of instructional guidance. Studies were coded as providing either low or high instructional guidance. For instance, in a study done by Hinzman (1997), students worked in groups without instructions on how to use the manipulatives provided to them. The only instruction provided to students was the objective of the lesson, to represent equations using different manipulatives such as colored disks and plastic cups. In contrast, high instructional guidance was provided within Getgood's (2000) study that explicitly taught students the concepts of greatest common factor and least common multiple with factor blocks.

Mathematical topic. The math topic of instruction was coded as a categorical variable with the following five categories: place value, arithmetic, geometry, fractions, and algebra. The category of arithmetic was used as a broad term encompassing math operations such as addition, subtraction, multiplication, and division. These categories were developed to be exhaustive of all topics present in the current body of literature.

Group versus individual instruction. Implementation of manipulatives strategies was coded as a dichotomous variable with studies being implemented at either the individual or the group level. Studies that were conducted in both small groups and whole class instruction were coded as group level. Bring (1972) presented an example of a study implemented at the individual level. In the study students were removed from the classroom and asked to individually complete a series of tasks or worksheets. As examples of classroom-level implementation, Slaughter (1980) and Battle (2007) examined the use of manipulatives with students who were provided instruction in a whole group classroom setting.

Instructional time. The time of treatment implementation was coded in days and then broken into tertiles to represent treatment times that were short (less than or equal to 14 days), medium (15–45 days), or long (greater than or equal to 46 days) in length. Several studies failed to report the duration of treatment. These studies were not included in the instructional time moderator analyses.

Outcome measure. The dependent variables from each study were coded into the following four classifications: retention, problem solving, transfer, and justification. Retention was defined as an outcome that required students to solve basic facts (for example from present sample, see, e.g., Smith & Montani, 2008). Problem solving included tasks in which students were not explicitly instructed on how to complete the assessment (see, e.g., LeBlanc, 1968; Shoecraft, 1971). Studies were classified as having a transfer outcome when students were asked to extend their knowledge to a new situation; for example, extending learned concepts of addition to multiplication (see, e.g., Ekman, 1967; Moody, Abell, & Bausell, 1971). Justification outcomes included activities in which participants were asked to provide explanations for why they used a given method to solve a problem (see, e.g., King, 1976).

Methodological Moderators

Peer review status. The publication source of each study was examined as a proxy variable for the quality of the study. Peer review status was coded as a dichotomous variable (published, not published). Studies were identified as published when they were located within a peer-reviewed journal. Other publication types such as thesis and dissertation studies found in research indexes (e.g., ProQuest) were coded as unpublished.

Research design. The design of each study was coded as either within-subjects, quasi-experimental, or experimental design. Studies classified as using within-subjects designs were those with a single group completing pre- and postassessments. The category of quasi-experimental designs included studies that manipulated the independent variable, but did not begin with random assignment to conditions. Studies with designs coded as experimental used random assignment to allocate individuals to conditions.

Implementer. Studies were also coded to identify who delivered treatment to students: a researcher or a teacher. When it was clear the person conducting the research was also the teacher of the students in the study, as was often the case in thesis and dissertation studies, the study was coded as treatment being delivered by the teacher.

Test type. Outcome measures were categorized as being researcher created or standardized assessments. For example, several studies used established assessments such as the Woodcock–Johnson (Woodcock & Johnson, 1989; see, e.g., Smith & Montani, 2008). Other studies used researcher-created materials, which were designed by the researcher specifically for use within the study (e.g., Hinzman, 1997).

Assumption of independence. The statistical assumption of independence was coded for each study. A dichotomous variable was created that distinguished studies that accounted for the nonindependence of observations that can occur when students in the sample are nested within classrooms from those studies that did not. Studies were coded by examining the degrees of freedom used in the analysis. For instance, Garcia (2004) implemented the intervention at the group level and used the classroom as the unit of analysis, which meets the assumption of independence. In contrast, Prigge (1978) implemented at the classroom level, but the degrees of freedom for the analysis indicated that the unit of analysis was the individual, which does not meet the statistical assumption of independence.

Analyses

The analysis plan included separate, but parallel, analyses for the aggregated and disaggregated data. Initially the effect sizes were examined in the aggregated data set. For the aggregated data, studies that reported multiple effect sizes were assigned a single effect size that was the average of the reported or calculated effect sizes (Rosenthal & Rubin, 1986). This procedure was followed to avoid the issue of nonindependence that can arise when multiple effects sizes are nested within a study, and to better address questions regarding the overall efficacy of the use of manipulatives on student learning. For the aggregated data, our procedure included the following steps. First, we calculated a weighted mean effect size across all studies. Next, we examined between-study variation in effect sizes using a Q statistic (Hedges, 1983). If statistically significant levels of between-study variation were found, we examined moderation of effect sizes based on both substantive and methodological features of the studies. All moderator variables were categorical.

A partitioning of variance approach (Hedges, 1982; Hedges & Olkin, 1985) was used to examine moderation. This approach uses a Q_{between} statistic to represent the between-group variability in effect sizes. This value can be referenced against a chi-square distribution with $k - 1$ degrees of freedom to test a null hypothesis of no difference in effect sizes across levels of the moderator. When differences were found, and there were more than two levels of the moderator variable, we conducted post hoc pairwise comparisons using an extension of the Scheffé procedure (see Hedges & Olkin, 1985) to maintain family-wise error rates at .05. As a last step in our analysis of the aggregated data, we computed a fail-safe N (Rosenthal, 1979) to assess the possible impact of studies with nonsignificant findings being overlooked in the analysis.

Examination of the disaggregated data was parallel to that of the analysis of the aggregated data. For these analyses, four sets of effect sizes were calculated according to the type of outcome used within the study. This approach allowed us to address more specific questions regarding the effect of manipulatives used for different outcomes while also avoiding any nonindependence among the effect sizes (no study reported multiple effect sizes within a single category of outcomes). For each of the four types of outcomes, we calculated an overall effect size and tested the level of between-study variation. When statistically significant levels of between-study variation were found, moderation analyses, as described above, were conducted.

Results

Coded characteristics of the 55 summarized studies are presented in Tables 2 and 3. Findings illustrate important differences among the studies examining the efficacy of math manipulatives. Of note are the following details: Fifty-five percent were published in a peer-reviewed scholarly journal; 46% of the studies were done after Sowell's 1989 meta-analysis; 56% of the studies were with third- and fourth-grade children; 76% of the studies were quasi-experimental or pre- and postdesigns; 75% failed to account for the statistical assumption of independence; and in 73% of the studies classroom teachers implemented the intervention.

Aggregated Data

The aggregated mean effect size of 0.37 was statistically significant ($p < .001$, 95% CI [0.30, 0.44]). Hedges's homogeneity test for effect sizes was also statistically significant, $Q(54) = 277.8$, $p < .001$, suggesting that between-study variation in effect sizes exceeded what would be expected by sampling error alone.

Moderator analysis. Table 4 summarizes the findings from the analysis of the moderator variables. The effect of mathematical topic was found to be statistically significant, $Q(4) = 29.8$, $p < .001$. Post hoc comparisons indicated that the mean effect size for fractions ($d = 0.69$) was statistically greater than that for arithmetic ($d = 0.27$) and algebra ($d = 0.21$). Instructional guidance was also a statistically significant moderator, $Q(1) = 6.3$, $p = .01$, with the effect size of studies with high instructional guidance ($d = 0.46$) greater than those with low guidance ($d = 0.29$). In addition, developmental status was statistically significant, $Q(2) = 11.7$, $p = .002$; samples consisting of children assumed to be concrete operational had a greater mean effect size ($d = 0.45$) than those with samples consisting of participants assumed to be in formal operations ($d = 0.16$). The effect sizes for studies of preoperational ($d = 0.33$) students was not significantly different from the effect sizes for studies of either concrete or formal operational students. Lastly, the instructional variable of time was significant, $Q(2) = 9.8$, $p = .008$. Post hoc comparisons indicated that the mean effect sizes for instruction provided for short lengths of time (≤14 days) and medium lengths of time (15–45 days) were not statistically significantly different ($d = 0.34$ and $d = 0.45$, respectively). However, both mean effect sizes were significantly greater than the mean effect for long lengths (≥46 days; $d = 0.14$).

Several moderators based on methodological characteristics of studies were also statistically significant. Test type, $Q(1) = 3.8$, $p = .05$, was statistically significant, with studies using standardized assessments having a higher mean effect size ($d = 0.49$) than researcher-created assessments of learning ($d = 0.33$). The statistical assumption of independence, $Q(1) = 5.6$, was also statistically significant ($p = .01$), with studies that met the statistical

Table 2
Descriptive Statistics for Continuous Variables

Variable	M	SD	Range	
			Minimum	Maximum
Age	9.8 years	2.4	5.5	17.0
Treatment time	25.0 days	42.7	1.0	180.0

Table 3
Descriptive Statistics for Categorical Variables

Variable	Category	%
Peer review status	Not published	55.3
	Published	44.6
Instructional guidance	Low	44.6
	High	55.3
Independence	Not met	75.0
	Met	25.0
Test type	Researcher created	73.3
	Standardized assessment	26.7
Implementer	Teacher	73.3
	Researcher	26.7
Research design	Within subjects	23.2
	Quasi-experimental	53.5
	Experimental	23.2
Mathematical topic	Arithmetic	42.8
	Place value	5.4
	Geometry	10.7
	Fractions	23.2
	Algebra	17.8
Perceptual richness	Yes	53.0
	No	47.0
Group vs. individual instruction	Individual	7.1
	Group	92.9
Outcome measures	Retention	94.6
	Transfer	24.5
	Justification	3.7
	Problem solving	16.0

assumption having a smaller mean effect size ($d = 0.19$) than studies that did not meet this assumption ($d = 0.41$). Additionally, study design was statistically significant, $Q(2) = 91.5$, $p < .001$; post hoc comparisons revealed that within-subjects studies had a higher mean effect size ($d = 1.22$) than quasi-experimental studies ($d = 0.28$) or studies using the experimental design ($d = 0.16$). No statistical difference between quasi-experimental and experimental designs was observed. Finally, peer-reviewed status, $Q(1) = 5.7$, $p = .01$, was statistically significant, with published studies having a greater mean effect size ($d = 0.46$) than unpublished studies ($d = 0.30$).

Disaggregated Data

Analysis of the disaggregated effect sizes by the learning outcomes revealed a mean effect size for retention of 0.59 (95% CI [0.52, 0.65]), whereas the mean effect size for problem solving was 0.46 (95% CI [0.23, 0.68], with both being statistically significant ($p < .001$). The mean effect size for transfer was 0.13 (95% CI [0.02, 0.23]), and the mean effect size for justification was 0.38 (95% CI [0.06, 0.70], both $p < .01$). For three outcomes, examination of between-study variance revealed variation in effect sizes exceeded what would be expected by sampling error: retention, $Q(52) = 719.2$; transfer, $Q(12) = 56.7$; and problem solving, $Q(8) = 56.3$, all $p < .001$. For justification the Q statistic was found to be nonsignificant, $Q(1) = 2.93$, $p = .23$. Moderator analyses were performed only for the three outcome measures with statistically significant variation in effect sizes.

Retention. Table 5 summarizes the findings for each of the coded moderator variables within the learning outcome of retention. As with the aggregated data, variation among effect sizes was impacted by both methodological and instructional variables. For instructional variables of interest, patterns similar to those found for the aggregated data emerged. Level of instructional guidance, $Q(1) = 106.5$, $p < .001$, was statistically significant, with high instructional guidance having a greater mean effect size ($d = 0.90$) than low instructional guidance ($d = 0.19$). Math topic was a statistically significant moderator, $Q(4) = 44.5$, $p < .001$, with studies concerning fractions ($d = 0.93$) and algebra ($d = 0.84$) having statistically higher mean effect sizes than studies teaching arithmetic ($d = 0.39$). Developmental status was significant, $Q(2) = 106.8$, $p < .001$, with post hoc comparisons revealing significant differences between all pairings of studies using preoperational ($d = -0.09$), concrete operational ($d = 0.81$), and formal operational ($d = 0.31$) samples. Perceptual richness was a significant moderator, $Q(1) = 36.4$, $p < .001$, with studies that used perceptually rich objects ($d = 0.28$) having a lower mean effect size than studies with bland or nondescript objects ($d = 0.77$). Instructional time was also a significant moderator, $Q(2) = 7.4$, $p = .02$. Post hoc comparisons revealed that that the mean effect size for instruction provided for short lengths of time ($d = 0\ .59$) was significantly greater than studies coded as medium lengths of time ($d = 0.35$), but not statistically different from long lengths of time ($d = 0.49$). Additionally, the difference between medium and long lengths of instructional time was not significant.

For methodological characteristics, several moderators were statistically significant. Peer review status was significant with published studies having a greater mean effect size ($d = 0.97$) than those that were unpublished ($d = 0.30$), $Q(1) = 90.7$, $p < .001$. Research design was significant, $Q(2) = 183.1$, $p < .001$. Studies using within-subjects designs had greater mean effect size ($d = 1.69$) than those using either quasi-experimental ($d = 0.35$) or experimental ($d = 0.47$) designs. The difference between studies using quasi-experimental and experimental designs was nonsignificant. Treatment implementer moderated effect size, $Q(1) = 89.1$, $p < .001$, with researcher-implemented treatments ($d = 0.13$) having a smaller mean effect size than teacher-implemented treatments ($d = 0.82$).

Problem solving. Table 6 contains the findings for the coded variables for studies measuring problem solving. Level of instructional guidance had a significant effect, $Q(1) = 19.1$, $p < .001$. High instructional guidance ($d = 1.06$) studies had a greater mean effect size than low instructional guidance studies ($d = 0.04$). Math topic also was a significant moderator, with $Q(3) = 45.1$, $p < .001$. Studies examining manipulatives with fractions had a greater mean effect size ($d = 2.50$) than those examining arithmetic ($d = 0\ .02$), place value ($d = 0\ .48$), and geometry ($d = 0.72$). The differences among the latter three were nonsignificant. Perceptual richness was significant, $Q(1) = 15.3$, $p < .001$, with studies that used perceptually rich objects ($d = -0.27$) having a lower mean effect size than studies with bland or nondescript objects ($d = 0.80$). Lastly, for instructional variables, time was a significant moderator, $Q(2) = 22.7$, $p < .001$, with instructional time coded as short ($d = 0.86$) and long ($d = 0.25$) in length having a greater mean effect size than studies coded as medium in length ($d = -0.62$). However, the difference between studies that were coded as short and long was not statistically significant.

Significant moderators related to research methods were present as well. Peer review status was significant, with published studies having a greater mean effect size ($d = 0.76$) than

Table 4
Variability of Effect Sizes Within Aggregated Data

Moderator	*k*	*N*	*d*	95% CI	$Q_{between}$
			Methodology characteristics		
Peer review status					5.7, $p = .01$
Published	24	4,190	0.46	[0.36, 0.56]	
Not published	31	3,047	0.30	[0.22, 0.39]	
Design					91.5, $p < .001$
Within subject	12	748	1.22_a	[1.03, 1.32]	
Quasi-experimental	30	5,173	0.28_b	[0.20, 0.37]	
Experimental	13	1,346	0.16_b	[0.02, 0.30]	
Implementer					1.6, $p = .19$
Researcher	15	1,285	0.29	[0.14, 0.43]	
Teacher	40	5,952	0.39	[0.32, 0.47]	
Test type					3.8, $p = .05$
Standardized	15	1,172	0.49	[0.35, 0.63]	
Researcher created	40	6,065	0.33	[0.26, 0.41]	
Independence					5.6, $p = .01$
Met	14	1,027	0.19	[0.03, 0.35]	
Not met	41	6,210	0.41	[0.33, 0.48]	
			Instructional characteristics		
Instructional guidance					6.3, $p = .01$
High	30	4,275	0.46	[0.36, 0.56]	
Low	25	2,962	0.29	[0.20, 0.38]	
Mathematical topic					29.8, $p < .001$
Place value	3	101	0.58_{ab}	[0.20, 0.96]	
Arithmetic	24	2,309	0.27_a	[0.16, 0.38]	
Geometry	6	662	0.37_a	[0.19, 0.56]	
Fractions	12	2,876	0.69_b	[0.55, 0.84]	
Algebra	10	1,348	0.21_a	[0.07, 0.34]	
Perceptual richness					0.21, $p = .64$
Yes	26	4,050	0.36	[0.27, 0.45]	
No	24	1,923	0.39	[0.28, 0.50]	
Group vs. individual					2.9, $p = .08$
Individual	4	388	0.58	[0.33, 0.83]	
Group	51	6,849	0.35	[0.29, 0.42]	
Development status					11.7, $p = .002$
Preoperational	10	1,256	0.33_a	[0.16, 0.49]	
Concrete	38	5,657	0.45_a	[0.37, 0.53]	
Formal	7	324	0.16_b	[0.01, 0.31]	
Instructional time					9.4, $p = .008$
≤14 days	27	4,340	0.34_a	[0.24, 0.45]	
15–45 days	16	1,353	0.45_a	[0.34, 0.57]	
≥46 days	6	540	0.14_b	[−0.01, 0.30]	

Note. For moderators with more than two levels, mean effect sizes with different subscripts are statistically different from one another, based on a family-wise Type I error probability of .05. CI = confidence interval.

those that were unpublished ($d = -0.33$), $Q(1) = 18.5$, $p < .001$. Design was also statistically significant, with $Q(2) = 26.3$, $p < .001$. Within-subject designs ($d = 1.23$) had higher mean effect sizes than quasi-experimental ($d = 0.27$) and experimental designs ($d = -0.08$). Additionally, implementer was significant, $Q(1) = 23.2$, $p < .001$; teacher-implemented treatments ($d = 0.82$) had greater effect sizes than researcher-delivered programs ($d = -0.39$). The assumption of statistical independence was significant, $Q(1) = 5.5$, $p = .01$. Studies that violated the assumption had greater effect sizes ($d = 0.61$) than those that met the assumption ($d = -0.01$).

Transfer. Table 7 contains the findings for each of the coded moderator variables for the learning outcome of transfer. Level of instructional guidance, $Q(1) = 6.7$, $p = .009$, was statistically significant. In contrast to the findings from the other outcomes, low levels of guidance produced a larger mean effect size ($d = 0.27$) than high levels of guidance ($d = 0.00$). Perceptual richness of the manipulative was also statistically significant, $Q(1) = 12.2$, $p < .001$. Again, differing from the results from the other outcomes, perceptually rich manipulatives had a higher mean effect size ($d = 0.48$) than bland manipulatives ($d = -0.02$) on transfer outcomes.

Two methodological variables were found to significantly moderate effect sizes: the design of the research, $Q(2) = 27.5$, $p < .001$, and the statistical assumption of independence, $Q(1) = 29.3$, $p < .001$. Unlike previous results, significant differences existed between studies in favor of experimental design, with experiments ($d = 0.40$) having a statistically significant higher mean effect size than both within-subjects ($d = 0.25$) and quasi-experiments ($d = -0.21$). Studies that met the assumption of independence had a smaller (and negative) mean effect size ($d = -0.67$) than those that did not meet the assumption ($d = 0.27$).

Table 5
Variability of Effect Sizes Within Retention

Moderator	*k*	*N*	*d*	95% CI	$Q_{between}$
Methodology characteristics					
Peer review status					90.7, $p < .001$
Published	23	4,162	0.97	[0.86, 1.07]	
Not published	30	2,978	0.30	[0.21, 0.39]	
Design					183.1, $p < .001$
Within subject	12	707	1.69_a	[1.52, 1.87]	
Quasi-experimental	29	5,187	0.35_b	[0.27, 0.44]	
Experimental	12	1,246	0.47_b	[0.34, 0.60]	
Implementer					89.1, $p < .001$
Researcher	15	6,135	0.13	[0.01, 0.24]	
Teacher	38	1,005	0.82	[0.73, 0.90]	
Test type					6.9, $p = .008$
Standardized	14	1,206	0.77	[0.62, 0.92]	
Researcher created	39	5,934	0.54	[0.47, 0.62]	
Independence					11.9, $p < .001$
Met	13	937	0.79	[0.66, 0.93]	
Not met	40	6,203	0.52	[0.44, 0.60]	
Instructional characteristics					
Instructional guidance					106.5, $p < .001$
High	31	4,531	0.90	[0.81, 0.99]	
Low	22	2,609	0.19	[0.08, 0.29]	
Mathematical topic					44.5, $p < .001$
Place value	3	101	0.70_{ab}	[0.37, 1.04]	
Arithmetic	25	2,387	0.39_b	[0.29, 0.48]	
Geometry	5	602	0.57_{ab}	[0.37, 0.78]	
Fractions	13	2,762	0.93_a	[0.78, 1.08]	
Algebra	9	1,288	0.84_a	[0.65, 1.03]	
Perceptual richness					36.4, $p < .001$
Yes	12	1,395	0.28	[0.14, 0.41]	
No	34	4,723	0.77	[0.69, 0.85]	
Group vs. individual					0.02, $p = .87$
Individual	5	481	0.57	[0.31, 0.82]	
Group	48	6,659	0.59	[0.52, 0.66]	
Development status					106.8, $p < .001$
Preoperational	8	707	-0.09_a	[−0.26, 0.07]	
Concrete	40	6,109	0.81_b	[0.73, 0.89]	
Formal	5	324	0.31_c	[0.10, 0.52]	
Instructional time					7.4, $p = .02$
≤14 days	25	3,133	0.59_a	[0.49, 0.69]	
15–45 days	15	1,261	0.35_b	[0.21, 0.49]	
≥46 days	7	952	0.49_{ab}	[0.28, 0.71]	

Note. For moderators with more than two levels, mean effect sizes with different subscripts are statistically different from one another, based on a family-wise Type I error probability of .05. CI = confidence interval.

Publication Bias

Publication bias refers to the possibility that results from studies showing statistically significant effects in the expected direction are more likely to be published than results form studies not showing such effects. This is commonly referred to as the file-drawer phenomenon (Rosenthal, 1979). To assess the possible impact of such bias, we included both published and unpublished manuscripts in the present review. However, given that our unpublished studies consist primarily of dissertation and thesis projects, we cannot rule out the possibility that other studies with nonsignificant findings have been excluded. Therefore, we conducted an analysis of publication bias to assess the potential impact of missing studies on meta-analytic results. Rosenthal's (1979) fail-safe *N* was calculated to determine how many studies with a null effect would be needed to attenuate the overall effect size to nonsignificance. This analysis revealed that approximately 9,501 studies would be needed to decrease the average effect size of manipulatives to nonsignificance.

Discussion

The purpose of this meta-analytic review was to examine the effect of using concrete manipulatives for teaching mathematics when compared with abstract symbolic instructional conditions. Additional research exploring comparisons of manipulatives against iconic representations of manipulatives would greatly enhance our understanding of instructional strategies that use concrete representations of mathematics. Currently, the findings from the present review of manipulatives suggest a small- to moderate-sized effect in favor of instructional strategies that use manipulatives when compared with abstract symbolic instruction. However,

Table 6
Variability of Effect Sizes Within Problem Solving

Moderator	*k*	*N*	*d*	95% CI	$Q_{between}$
		Methodology characteristics			
Peer review status					18.5, $p < .001$
Published	7	335	0.76	[0.50, 1.03]	
Not published	2	142	−0.33	[−0.07, 0.09]	
Design					26.3, $p < .001$
Within subject	3	146	1.23_a	[−0.85, 1.60]	
Quasi-experimental	3	144	0.27_b	[−0.20, 0.75]	
Experimental	3	187	-0.08_b	[−0.43, 0.25]	
Implementer					23.2, $p < .001$
Researcher	2	116	−0.39	[−1.85, 0.02]	
Teacher	7	361	0.82	[0.55, 1.09]	
Test type					8.2, $p = .002$
Standardized	3	201	−0.03	[−0.43, 0.36]	
Researcher created	6	276	0.69	[0.42, 0.97]	
Independence					5.5, $p = .01$
Met	2	106	−0.01	[−0.47, 0.43]	
Not met	7	371	0.61	[0.35, 0.87]	
		Instructional characteristics			
Instructional guidance					19.1, $p < .001$
High	5	202	1.06	[0.71, 1.42]	
Low	4	275	0.04	[−1.30, 1.39]	
Mathematical topic					45.1, $p < .001$
Place value	1	24	0.48_a	[−0.32, 1.29]	
Arithmetic	5	246	0.02_a	[−0.26, 0.30]	
Geometry	1	93	0.72_a	[0.20, 1.24]	
Fractions	2	114	2.50_b	[1.82, 3.18]	
Algebra	0				
Perceptual richness					15.3, $p < .001$
Yes	2	100	−0.27	[−0.72, 0.18]	
No	6	377	0.80	[0.52, 1.08]	
Group vs. individual					
Individual	0				
Group	9	477	0.46	[0.23, 0.68]	
Development status					2.2, $p = .33$
Preoperational	1	66	0.08	[−0.58, 0.75]	
Concrete	7	318	0.45	[0.18, 0.72]	
Formal	1	93	0.72	[0.20, 1.24]	
Instructional time					22.7, $p < .001$
≤14 days	6	290	0.86_a	[0.56, 1.16]	
15–45 days	1	76	-0.62_b	[−1.17, −0.62]	
≥46 days	2	111	0.25_a	[−0.19, 0.69]	

Note. For moderators with more than two levels, mean effect sizes with different subscripts are statistically different from one another, based on a family-wise Type I error probability of .05. CI = confidence interval.

these results cannot be used as evidence that manipulatives are beneficial for learning when making comparisons to other mathematic instructional strategies. Furthermore, an examination of effect sizes across instructional characteristics and learning outcomes revealed that the effectiveness of manipulatives is complex and requires consideration of instructional characteristics and learning outcomes.

Concrete manipulatives have been proposed as an effective strategy in aiding students in problem solving and transfer of mathematical understanding (Burns, 1996; NCTM, 2000). Effect sizes when separated by learning outcome did not follow this assumed pattern. In fact, instruction that used manipulatives produced a moderate- to large-sized effect when students were measured on retention and small effects when higher level outcomes such as problem solving, transfer, and justification were considered. Taken as a whole, the aggregated and disaggregated findings indicate that concrete manipulatives may have a differential impact on learning outcomes. These differential outcomes should be considered for future empirical examination and when teaching mathematics concepts with manipulatives.

Instructional Characteristics

Level of instructional guidance, mathematical topic, development status, perceptual richness, and instructional time were statistically significant moderators of the effects of using concrete manipulatives. The moderators of developmental status, level of instructional guidance, type of manipulative, and instructional time are of particular interest due to the connections between these moderators and current developmental and instructional theories.

Table 7
Variability of Effect Sizes Within Transfer

Moderator	*k*	*N*	*d*	95% CI	Q_{between}
		Methodology characteristics			
Peer review status					2.2, $p = .13$
Published	8	2,871	0.19	[0.05, 0.32]	
Not published	5	582	0.04	[−0.15, 0.25]	
Design					27.5, $p < .001$
Within subject	2	72	0.25_{ab}	[0.69, 0.12]	
Quasi-experimental	6	702	-0.21_{a}	[−0.39, −0.03]	
Experimental	5	2,415	0.40_{b}	[0.25, 0.56]	
Implementer					1.4, $p = .23$
Researcher	4	259	0.06	[−0.20, 0.33]	
Teacher	9	3,194	0.16	[0.04, 0.29]	
Test type					0.92, $p = .33$
Standardized	1	60	0.18	[−0.43, 0.80]	
Researcher created	12	3,393	0.14	[0.03, 0.26]	
Independence					29.3, $p < .001$
Met	3	225	−0.61	[−0.91, −0.31]	
Not met	10	3,228	0.27	[0.15, 0.39]	
		Instructional characteristics			
Instructional guidance					6.7, $p = .009$
High	8	2,385	0.00	[−0.16, 0.16]	
Low	5	1,068	0.27	[0.12, 0.43]	
Mathematical topic					1.5, $p = .83$
Place value	1	12	0.33	[−0.28, 0.96]	
Arithmetic	7	757	0.16	[0.00, 0.36]	
Geometry	1	147	0.24	[−0.15, 0.65]	
Fractions	3	1,996	0.09	[−0.14, 0.32]	
Algebra	1	541	0.09	[−0.18, 0.36]	
Perceptual richness					12.2, $p < .001$
Yes	5	1,359	0.48	[0.25, 0.62]	
No	6	2,094	−0.02	[−0.23, 0.17]	
Group vs. Individual					
Individual					
Group					
Developmental Status					
Preoperational					
Concrete	13	3,453	0.14	[0.03, 0.26]	
Formal					
Instructional time					3.6, $p = .06$
≤14 days	7	1,625	0.03	[−0.13, 0.21]	
15–45 days	6	1,828	0.22	[0.07, 0.37]	
≥46 days					

Note. For moderators with more than two levels, mean effect sizes with different subscripts are statistically different from one another, based on a family-wise Type I error probability of .05. CI = confidence interval.

Developmental status. Several contemporary theorists propose that ability to reason abstractly is the pinnacle of cognitive development. According to these theorists, concrete manipulatives should be provided when younger learners begin studying abstract mathematical concepts (Bruner, 1964; Piaget, 1962). Within this developmental framework, it is expected that providing manipulatives allows educators to represent abstract concepts with concrete representations. These concrete representations are expected to facilitate the construction of meaning for pre- and concrete operational children and result in positive cognitive consequences. In contrast, students who are assumed to have reached formal operations are not expected to derive comparable cognitive benefits from the provision of concrete manipulatives.

Our findings provide partial support for these developmental predictions. At the aggregated level, studies that included children assumed to have facility with concrete operations showed medium to large effect sizes, whereas studies comprising formal operational students had relatively smaller effect sizes. Within the learning outcome of retention, the pattern of manipulatives being the most efficacious for students within the assumed concrete operations stage remains; however, studies consisting of samples of preoperational-age children revealed a statistically lower and negative mean effect size ($d = -0.09$) than studies consisting of samples of assumedly concrete or formal operational students. Recently, developmental theorists have proposed explanations for why concrete manipulatives may be less effective with younger children. According to these explanations, concrete manipulatives may not be as effective with younger children because they may struggle with the concept that an object can stand for the item while simultaneously representing a larger mathematical concept (DeLoache, 2000; Uttal et al., 2009). Future research should examine how concrete manipulatives assist in developing foundational mathematical concepts with younger children.

Instructional guidance. Conflicting recommendations have been provided to practitioners concerning the level of instructional guidance offered to students during the learning process. Those who recommend high levels of instructional guidance propose that instructional guidance provides students with explicit opportunities to select pertinent information, organize the information into coherent structures, and incorporate the new information with prior knowledge (Mayer, 2003). According to this model, low levels of instructional guidance do not promote this process due to the lack of explicit guidance selecting relevant information. Results from the aggregated, retention, and problem-solving data support this model, with high levels of guidance being associated with higher levels of student learning. This finding also aligns with prior research examining the efficacy of manipulatives in listening and reading instruction (Glenberg, Brown, & Levin, 2007; Glenberg, Jaworski, Rischal, & Levin, 2007; Marley et al., 2007). These researchers have suggested that the scripted manipulation of objects helps students establish connections between concrete representations and their abstract referents (i.e., words), which in turn enhances comprehension.

Proponents of low instructional guidance contend that students who reach proficiency with limited or no instructional guidance develop greater conceptual understandings and are subsequently more adept at transferring this knowledge to novel circumstances (Schauble, 1996; Stohr-Hunt, 1996). Martin's (2009) theory of physically distributed learning supports this notion with the explanation that students are able to impose their own meanings on manipulatives. This development of self-relevant meaning allows for greater flexibility and for learning to be transferred to novel circumstances. The moderator analysis associated with the transfer of learning outcome provides partial support for this perspective, with low instructional guidance studies having larger effect sizes on transfer of learning relative to high levels of guidance. Future research is needed to better understand what level of instructional guidance is optimum for student learning with manipulatives. More specifically, research is needed to examine how level of instructional guidance may need to vary depending upon learning objective.

Perceptual richness. The perceptual richness of manipulatives has been identified as a potential deterrent to student learning and performance (Kaminski et al., 2009; Martin & Schwartz, 2005; McNeil et al., 2009). Superficial details that are present in perceptually rich manipulatives have been shown to distract children when asked to perform a math word problem, resulting in students making more errors in solving the math problem but proportionally fewer conceptual errors than students who used bland manipulatives (McNeil et al., 2009).

The moderator analysis within the retention and problem-solving data provides additional support for the idea that perceptually rich manipulatives suppress student learning. Within the learning outcome of retention, studies utilizing perceptually rich manipulatives had a smaller effect on student measures of immediate performance. Additionally, specific to the learning outcome of problem solving, those studies that used perceptually rich manipulatives revealed a lower and negative mean effect size ($d = -0.27$) when compared with studies that used bland manipulatives during the problem solving process.

Results on transfer of learning, an outcome that requires greater conceptual understanding of the mathematics concepts, indicated that perceptually rich manipulatives may enhance student learning. With findings indicating that studies that used bland manipulatives had a lower and negative mean effect size ($d = -0.02$) than studies using perceptually rich manipulatives. However, it is important to note that this finding contradicts previous cognitive research that suggests that the perceptual richness of images inhibits the transfer of learning (Goldstone & Sakamoto, 2003; Kaminski, Sloutsky, & Heckler, 2008). To better understand this finding, additional research examining the relationship between the perceptual richness of manipulatives and their effects on a variety of student learning outcomes is warranted.

Instructional time. The length of instructional time provided to students has been established as an essential variable to learning (Rosenshine & Berliner, 1978). Furthermore, experiments contrasting direct and discovery learning as means to improving inquiry skills have implicated length of instructional time as an explanation for inconsistent empirical results (Dean & Kuhn, 2007) within the instructional strategy literature. According to Dean and Kuhn (2007), in order for student-controlled strategies to be effective, students must engage in instruction over an extended period. In addition, Sowell's (1989) meta-analysis provides evidence that extended use of manipulatives had a positive effect on measures of retention. Results from the moderator analysis of the present study contradict these findings. Studies that were less than 45 days had a higher mean effect on student learning within the aggregated data. Additionally, within the learning outcomes of retention, studies that were less than 14 days had a higher mean effect than studies that were longer (i.e., more than 15 days), and within problem solving, studies that were coded at being medium in length (15 to 45 days) had a lower and negative mean effect size ($d = -0.62$) than short or long studies. A possible explanation for these contradictory findings was the inability of the coding to disentangle instructional time with the length of study. Therefore, further research that specifically examines varying lengths of instructional time with manipulatives is needed to provide a better understanding of how this instructional variable moderates the overall effectiveness of manipulatives.

Methodological Characteristics

Methodological aspects of studies presented additional variation in the overall findings. This variation in effect sizes due to methodological characteristics is related to the validity of the inferences made from the results of this literature. Methodological characteristics that significantly moderated effect sizes included peer review status, research design, implementer, type of test, and accounting for the statistical assumption of independence. Moderation within these variables raises specific concerns related to statistical conclusion validity and internal validity.

Statistical conclusion validity. Validity concerning the inferences made about the covariation between treatment and outcome is of specific concern when examining the efficacy of teaching strategies. Threats to statistical conclusion validity may lead to results that either overestimate or underestimate the magnitude of the covariation between the treatment and the outcome (Shadish et al., 2002). When threats of this nature have a high prevalence in the literature, conclusions regarding the effectiveness of the teaching strategy are limited. This concern is supported by the findings that a greater effect size was produced when statistical independence is not accounted for in the analysis. Furthermore, published studies and within-subjects studies produced larger effect sizes. The latter suggests that the results of within-subjects studies should be carefully evaluated. The former could be an indication that published studies are of higher quality and produce significant results, or that significant findings tend to be published over findings that fail to reach statistical significance.

Internal validity. The ability to make an inference about the causal effect of manipulatives on student learning is grounded in research design. At a minimum, to build a case for strong internal validity, studies must be able to account for plausible rival explanations. The most effective way to control for threats to internal validity is the use of an experimental design. By randomly assigning participants to conditions, plausible threats to internal validity are minimized. The finding that studies that did not use random assignment had a significantly higher effect size than those from experimental designs emphasizes the need for researchers and practitioners to be cautious when making prescriptive statements concerning the efficacy of manipulatives.

Limitations

Limitations of meta-analyses are often similar to those for primary studies (Card, 2012). Three such limitations common to both meta-analytical and primary studies that are relevant to the current meta-analysis are the potential problem of generalizing results, issues related to incomplete or missing data, and the difficultly of drawing sound inference about causal effects.

Specific to this meta-analysis is the possible limitation of generalizability due to our definitions of manipulative and control groups. For example, studies that used a scale were not included in the sample, nor were studies that made comparisons between manipulatives and pictures. Inclusion of studies that used tools or pictures may alter conclusions regarding mathematics manipulatives. Therefore, the circumscribed conclusion that can be drawn is that a positive effect of manipulatives was found on student learning outcomes when it was compared with conditions in which no manipulatives or other concrete materials (e.g., pictures) were used and that the relationship between manipulatives and learning within this comparison is moderated by different instructional characteristics. Further research should be conducted to collect evidence comparing manipulatives to other learning strategies.

Another potential limitation on generalizability is posed by the previously mentioned file-drawer problem. A careful attempt was made to exhaustively search for unpublished manuscripts; however, some studies examining the efficacy of manipulatives may not have been identified. This means the possibility of publication bias is still present. Additionally, in many instances, studies included in the meta-analysis did not report enough information for all variables of interest to be coded. The lack of information from individual studies created numerous occasions of missing values for the moderator variables. The inability to use all studies in each moderator analysis could have decreased the statistical power of the meta-analysis to detect differences that may exist. The lack of information from studies also created many missed opportunities to further develop an understanding of the efficacy of manipulatives. Detailed information related to the procedure of the study (i.e., interval of time between outcomes measures, the use of manipulatives at time of testing, and details related to actual activities of participants) may have yielded a slightly different overall outcome on student learning. For example, the lack of details provided in articles limited the ability to make distinctions between learning activities that were conducted within each study. Clarification of the activities conducted within each study may have provided an opportunity to use established frameworks such as Chi's (2009) taxonomy of learning activities.

Finally, as noted by Card (2012), the strength of conclusions from any meta-analysis is based on the quality of the research design for both the meta-analysis and each of the constituent studies. For example, results from a meta-analysis of high-quality studies will always yield better conclusions than a meta-analysis of low-quality studies. Therefore, the observed methodological differences in study quality may affect the quality of the present results. In addition, it should be emphasized that the reported moderator effects are based on the observed covariation between the coded moderator variables and the values of the study effect sizes. As such, moderator effects should not be construed as strong evidence for the causal effect of a moderator variable on the effectiveness of using manipulative.

Conclusion

Results from this meta-analysis begin to focus the inconsistencies seen within the manipulation-based literature. Findings indicate that using manipulatives in mathematics instruction produces a small- to medium-sized effect on student learning when compared with instruction that uses abstract symbols alone. Additionally, results revealed that the strength of this effect is dependent upon other instructional variables. Instructional variables such as the perceptual richness of an object, level of guidance offered to students during the learning process, and the development status of the learner moderate the efficacy of manipulatives. The finding that specific instructional variables either suppress or increase the efficacy of manipulatives suggests that simply incorporating manipulatives into mathematics instruction may not be enough to increase student achievement in mathematics. These contextual variables therefore must be considered when planning instruction. It is our hope that the results of this meta-analysis will further stimulate math manipulatives research and assist others in generating new and more specific hypotheses investigating this instructional strategy.

References

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

*Aburime, E. F. (2007). How manipulatives affect the mathematics achievement of students in Nigerian schools. *Education Research Quarterly, 3*(1), 3–16.

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology, 103*(1), 1–18. doi:10.1037/a0021017

*Anderson, G. R. (1957). Visual–tactual devices and their efficacy: An experiment in grade eight. *Arithmetic Teacher, 4*(3), 196–203.

*Aurich, M. R. (1963). *A comparative study to determine the effectiveness of the Cuisenaire method of arithmetic instruction with children at the* *first grade level* (Unpublished doctoral dissertation). Catholic University of America, Washington, DC.

*Babb, J. H. (1975). The effects of textbook instruction, manipulatives and imagery on recall of the basic multiplication facts. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 36,* 4378.

Baranes, R., Perry, M., & Stigler, J. W. (1989). Activation of real-world knowledge in the solution of word problems. *Cognition and Instruction, 6*(4), 287–318. doi:10.1207/s1532690xci0604_1

*Battle, T. S. (2007). *Infusing math manipulatives: The key to an increase in academic achievement in the mathematics classroom.* (ERIC Document Reproduction Service No. ED498579)

Biazak, J. E., Marley, S. C., & Levin, J. R. (2010). Does an activity-based learning strategy improve preschool children's memory for narrative passages? *Early Childhood Research Quarterly, 25*(4), 515–526. doi:10.1016/j.ecresq.2010.03.006

Billstein, R., Libeskind, S., & Lott, J. W. (2009). *A problem solving approach to mathematics: For elementary school teachers*. Boston, MA: Addison Wesley.

*Bring, C. R. (1972). Effects of varying concrete activities on the achievement of objectives in metric and non-metric geometry by students of grades five and six. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 32*(7), 3775.

Brown, M. C., McNeil, N. M., & Glenberg, A. M. (2009). Using concreteness in education: Real problems, potential solutions. *Child Development Perspectives, 3*(3), 160–164. doi:10.1111/j.1750-8606.2009.00098.x

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review, 31,* 21–32.

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

Burns, M. (1996). How to make the most of math manipulatives. *Instructor, 105*(7), 45–51.

*Butler, F. M., Miller, S. P., Crehan, K., Babbitt, B., & Pierce, T. (2003). Fraction instruction for students with mathematics disabilities: Comparing two teaching sequences. *Learning Disabilities Research & Practice, 18*(2), 99–111. doi:10.1111/1540-5826.00066

Canobi, K. H. (2005). Children's profiles of addition and subtraction understanding. *Journal of Experimental Child Psychology, 92*(3), 220–246. doi:10.1016/j.jecp.2005.06.001

Carbonneau, K. J., & Marley, S. C. (in press). Activity-based learning strategies and academic achievement. In J. A. C. Hattie & E. M. Anderman (Eds.), *The international handbook of student achievement*. New York, NY: Routledge.

Card, N. A. (2012). *Applied meta-analysis for social science research*. New York, NY: Guilford Press.

*Carmody, L. M. (1970). A theoretical and experimental investigation into the role of concrete and semi-concrete materials in the teaching of elementary school mathematics. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 31,* 3407.

Chi, M. T. H. (2009). Active–constructive–interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1*(1), 73–105. doi:10.1111/j.1756-8765.2008.01005.x

Clark, J. M., & Paivio, A. (1987). A dual coding perspective on encoding processes. In M. A. McDaniel & M. Pressley (Eds.), *Imagery and related mnemonic processes: Theories, individual differences, and applications* (pp. 5–33). New York, NY: Springer-Verlag.

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

*Cook, D. M. (1967). *Research and development activities in R & I units of two elementary schools in Janesville, Wisconsin, 1966–1967 (Report No. TR-45)*. Washington, DC: U.S. Department of Health, Education and Welfare.

Copley, J. V. (2000). *The young child and mathematics*. Washington, DC: National Association for the Education of Young Children. (ERIC Document Reproduction Service No. ED448906)

*Cramer, K. A., Post, T. R., & delMas, R. C. (2002). Initial fraction learning by fourth- and fifth-grade students: A comparison of the effects of using commercial curricula with the effects of using the Rational Number Project Curriculum. *Journal for Research in Mathematics Education, 33*(2), 111–144. doi:10.2307/749646

*Dawson, D. T. (1955). An experimental approach to the division idea. *Arithmetic Teacher, 2*(1), 6–9.

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91*(3), 384–397. doi:10.1002/sce.20194

DeLoache, J. S. (2000). Dual representation and young children's use of scale models. *Child Development, 71*(2), 329–338. doi:10.1111/1467-8624.00148

*Dyer, L. (1996). *An investigation of the use of algebraic manipulatives with community college students* (Unpublished doctoral dissertation). University of Missouri, Columbia, MO.

*Egan, D. (1990). *The effects of using Cuisenaire rods on math achievement of second grade students* (Unpublished doctoral dissertation). Central Missouri State University, Warrensburg, MO.

*Ekman, L. G. (1967). A comparison of the effectiveness of different approaches to the teaching of addition and subtraction algorithms in the third grade. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 27,* 2275–2276.

Engelkamp, J., Zimmer, H. D., Mohr, G., & Sellen, O. (1994). Memory of self-performed tasks: Self-performing during recognition. *Memory & Cognition, 22*(1), 34–39. doi:10.3758/BF03202759

Fennema, E. H. (1972). The relative effectiveness of a symbolic and a concrete model in learning a selected mathematical principle. *Journal for Research in Mathematics Education, 3*(4), 233–238. doi:10.2307/748490

*Fujimura, N. (2001). Facilitating children's proportional reasoning: A model of reasoning processes and effects of intervention on strategy change. *Journal of Educational Psychology, 93*(3), 589–603. doi:10.1037/0022-0663.93.3.589

*Garcia, E. P. (2004). *Using manipulatives and visual cues with explicit vocabulary enhancement for mathematics instruction with grade three and four low achievers in bilingual classrooms* (Unpublished doctoral dissertation). Texas A&M University, College Station, TX.

*Getgood, J. F. (2000). *The effect of factor blocks, a manipulative, in student understanding of greatest common factor, least common multiple and prime factorization* (Unpublished doctoral dissertation). George Mason University, Fairfax, VA.

Glenberg, A. M., Brown, M., & Levin, J. R. (2007). Enhancing comprehension in small reading groups using a manipulation strategy. *Contemporary Educational Psychology, 32*(3), 389–399. doi:10.1016/j.cedpsych.2006.03.001

Glenberg, A. M., Gutierrez, T., Levin, J. R., Japuntich, S., & Kaschak, M. P. (2004). Activity and imagined activity can enhance young children's reading comprehension. *Journal of Educational Psychology, 96*(3), 424–436. doi:10.1037/0022-0663.96.3.424

Glenberg, A. M., Jaworski, B., Rischal, M., & Levin, J. (2007). What brains are for: Action, meaning, and reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 221–240). Mahwah, NJ: Erlbaum

*Goins, K. B. (2001). Comparing the effects of visual and algebra tile manipulative methods on student skill and understanding of polynomial multiplication. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 62,* 12.

Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology, 46*(4), 414–466. doi:10.1016/S0010-0285(02)00519-4

*Gürbüz, R. (2010). The effect of activity-based instruction on conceptual development of seventh grade students in probability. *International Journal of Mathematical Education in Science and Technology, 41*(6), 743–767. doi:10.1080/00207391003675158

*Hawkins, V. J. (1982). A comparison of two methods of instruction, a saturated learning environment and traditional learning environment: Its effects on achievement and retention among female adolescents in first-year algebra. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 43,* 02.

Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics, 7*(2), 119–137. doi:10.2307/1164961

Hedges, L. V. (1983). A random effects model for effect sizes. *Psychological Bulletin, 93*(2), 388–395. doi:10.1037/0033-2909.93.2.388

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

*Hiebert, J., Wearne, D., & Taber, S. (1991). Fourth graders' gradual construction of decimal fractions during instruction using different physical representations. *Elementary School Journal, 91*(4), 321–341. doi:10.1086/461658

*Hinzman, K. P. (1997). *Use of manipulatives in mathematics at the middle school level and their effects on student grades and attitudes (ERIC Document Reproduction Service No. ED411150).*

*Johnson, R. L. (1970). *Effects of varying concrete activities on achievement of objectives in perimeter, area and volume by students grades four five and six* (Unpublished doctoral dissertation). University of Colorado, Denver, CO.

*Jordan, L., Miller, M. D., & Mercer, C. D. (1999). The effects of concrete to semiconcrete to abstract instruction in the acquisition and retention of fraction concepts and skills. *Learning Disabilities, 9*(3), 115–122.

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. (2008). The advantage of abstract examples in learning math. *Science, 320*(5875), 454–455. doi:10.1126/science.1154659

Kaminski, J. A., Sloutsky, V. M., & Heckler, A. (2009). Transfer of mathematical knowledge: The portability of generic instantiations. *Child Development Perspectives, 3*(3), 151–155. doi:10.1111/j.1750-8606.2009.00096.x

King, V. C. (1976). A study of the effect of selected mathematics instructional sequences on retention and transfer with logical reasoning controlled. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 37*(5), 2696–2697.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. doi:10.1207/s15326985ep4102_1

Kormi-Nouri, R., Nyberg, L., & Nilsson, L. G. (1994). The effect of retrieval enactment on recall of subject-performed tasks and verbal tasks. *Memory & Cognition, 22*(6), 723–728. doi:10.3758/BF03209257

*Kuhfittig, P. K. F. (1974). The relative effectiveness of concrete aids in discovery learning. *School Science and Mathematics, 74*(2), 104–108. doi:10.1111/j.1949-8594.1974.tb09207.x

*LeBlanc, J. F. (1968). The performance of first grade children in four levels of conservation of numerousness and three IQ groups when solving arithmetic subtraction problems. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 29,* 67A.

Lefrançois, G. R. (1997). *Psychology for teachers*. Belmont, CA: Wadsworth.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*Lucas, J. S. (1966). The effects of attribute block training on children's development of arithmetic concepts. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 27,* 2400–2401.

*Lucow, W. H. (1964). An experiment with the Cuisenaire method in grade three. *American Educational Research Journal, 1*(3), 159–167. doi:10.2307/1162217

Marley, S. C., & Levin, J. R. (2006). Pictorial illustrations, visual imagery, and motor activity: Their instructional implications for Native American children with learning disabilities. In R. J. Morris (Ed.), *Disability research and policy: Current perspectives* (pp. 103–123). Mahwah, NJ: Erlbaum.

Marley, S. C., & Levin, J. R. (2011). When are prescriptive statements in educational research justified? *Educational Psychology Review, 23*(2), 197–206. doi:10.1007/s10648-011-9154-y

Marley, S. C., Levin, J. R., & Glenberg, A. M. (2007). Improving Native American children's listening comprehension through concrete representations. *Contemporary educational psychology, 32*(3), 537–550. doi:10.1016/j.cedpsych.2007.03.003

Marley, S. C., Levin, J. R., & Glenberg, A. M. (2010). What cognitive benefits does an activity-based reading strategy afford young Native American readers? *Journal of Experimental Education, 78*(3), 395–417. doi:10.1080/00220970903548061

Marley, S. C., & Szabo, Z. (2010). Improving children's listening comprehension with a manipulation strategy. *Journal of Educational Research, 103*(4), 227–238. doi:10.1080/00220670903383036

Marley, S. C., Szabo, Z., Levin, J. R., & Glenberg, A. M. (2011). Investigation of an activity-based text-processing strategy in mixed-age child dyads. *Journal of Experimental Education, 79*(3), 340–360. doi:10.1080/00220973.2010.483697

Martin, T. (2009). A theory of physically distributed learning: How external environments and internal states interact in mathematics learning. *Child Development Perspectives, 3*(3), 140–144. doi:10.1111/j.1750-8606.2009.00094.x

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science, 29*(4), 587–625. doi:10.1207/s15516709cog0000_15

Mayer, R. E. (2003). *Learning and instruction*. Upper Saddle River, NJ: Prentice Hall.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist, 59*(1), 14–19. doi:10.1037/0003-066X.59.1.14

*McClung, L. W. (1998). *A study on the use of manipulatives and their effect on student achievement in high school Algebra I class.* (ERIC Document Reproduction Service No. ED425077)

McNeil, N., & Jarvin, L. (2007). When theories don't add up: disentangling he manipulatives debate. *Theory Into Practice, 46*(4), 309–316. doi:10.1080/00405840701593899

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*(2), 171–184. doi:10.1016/j.learninstruc.2008.03.005

*Miller, J. W. (1964). An experimental comparison of two approaches to teaching multiplication of fractions. *Journal of Educational Research, 57*(9), 468–471.

Montessori, M. (1964). *The Montessori method*. New York, NY: Schocken.

*Moody, W. B., Abell, R., & Bausell, R. B. (1971). The effect of activity-oriented instruction upon original learning, transfer, and retention. *Journal for Research in Mathematics Education, 2*(3), 207–212. doi:10.2307/749045

*Nasca, D. (1966). Comparative merits of a manipulative approach to second-grade arithmetic. *Arithmetic Teacher, 13*(3), 221–225.

National Center for Education Statistics. (2008). *Highlights from TIMSS 2007: Mathematics and science achievement of U.S. fourth- and eighth-grade students in an international context*. Washington, DC: U.S. Department of Education. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2009001

National Center for Education Statistics. (2011). *The nation's report card: Mathematics 2011*. Washington, DC: U.S. Department of Education. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2012458

National Council of Teachers of Mathematics. (2000). *Standards & focal points*. Retrieved from http://www.nctm.org/standards/default.aspx?id=58

*Nichols, E. J. (1972). A comparison of two methods of instruction in multiplication and division for third grade pupils. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 32,* 6011.

*Nickel A. P. (1971). A multi-experience approach to conceptualization for the purpose of improvement of verbal problem-solving in arithmetic. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 32,* 2917–2918.

*Nishida, T. K. (2007). The use of manipulatives to support children's acquisition of abstract math concepts. *Dissertation Abstracts International: Section B. Sciences and Engineering, 69*(1), 718.

*Norman, M. (1955). *Three methods of teaching basic division facts* (Unpublished doctoral dissertation). State University of Iowa, Ames, IA.

*Olkun, S. (2003). Comparing computer versus concrete manipulatives in learning 2D geometry. *Journal of Computers in Mathematics and Science Teaching, 22*(1), 43–56.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.

*Paolini, M. W. (1977). The use of manipulative materials versus nonmanipulative materials in a kindergarten mathematics program. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 39,* 5382.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York, NY: Basic Books.

*Peterson, S. K., Mercer, C. D., & O'Shea, L. (1988). Teaching learning disabled students place value using the concrete to abstract sequence. *Learning Disabilities Research, 4*(1), 52–56.

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

Piaget, J., & Coltman, D. (1974). *Science of education and the psychology of the child*. New York, NY: Grossman.

*Prigge, G. R. (1978). The differential effects of the use of manipulative aids on the learning of geometric concepts by elementary school children. *Journal for Research in Mathematics Education, 9*(5), 361–367. doi:10.2307/748772

Resnick, L. B., & Omanson, S. F. (1987). Learning to understand arithmetic. In R. Glaser (Ed.), *Advances in instructional psychology* (Vol. *3,* pp. 41–95). Hillsdale, NJ: Erlbaum.

Rittle-Johnson, B., & Koedinger, K. R. (2005). Designing knowledge scaffolds to support mathematical problem solving. *Cognition and Instruction, 23*(3), 313–349. doi:10.1207/s1532690xci2303_1

*Robinson, E. B. (1978). *The effects of a concrete manipulative on attitude toward mathematics and levels of achievement and retention of a mathematical concept among elementary students* (Unpublished doctoral dissertation). East Texas State University, Commerce, TX.

Rosenshine, B. V., & Berliner, D. C. (1978). Academic engaged time. *British Journal of Teacher Education, 4*(1), 3–16. doi:10.1080/0260747780040102

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

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin, 99*(3), 400–406. doi:10.1037/0033-2909.99.3.400

Sarama, J., & Clements, D. H. (2009). "Concrete" computer manipulatives in mathematics education. *Child Development Perspectives, 3*(3), 145–150. doi:10.1111/j.1750-8606.2009.00095.x

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*(1), 102–119. doi:10.1037/0012-1649.32.1.102

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

*Sherman, J., & Bisanz, J. (2009). Equivalence in symbolic and nonsymbolic contexts: Benefits of solving problems with manipulatives. *Journal of Educational Psychology, 101*(1), 88–100. doi:10.1037/a0013156

*Shoecraft, P. J. (1971). *The effects of provisions for imagery through materials and drawings on translating algebra word problems.* (ERIC Document Reproduction Service No. ED071857)

*Slaughter, H. B. (1980, April). *Using Title 1 control group for evaluation research of a supplemental mathematics project for third and fifth grade students*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

*Smith, L. F., & Montani, T. O. (2008). The effects of instructional consistency: Using manipulatives and teaching strategies to support resource room mathematics instructions. *Learning Disabilities, 15*(2), 71–76.

Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education, 20*(5), 498–505. doi:10.2307/749423

*Steen, K., Brooks, D., & Lyon, T. (2006). The impact of virtual manipulatives on first grade geometry instruction and learning. *Journal of Computers in Mathematics and Science Teaching, 25*(4), 373–391.

*Steger, C. (1977). The effects of two classes of sensory stimuli and race on the acquisition and transfer of the mathematics principle place value. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 37,* 4120.

Stohr-Hunt, P. M. (1996). An analysis of frequency of hands-on experience and science achievement. *Journal of Research in Science Teaching, 33*(1), 101–109. doi:10.1002/(SICI)1098-2736(199601)33:1<101::AID-TEA6>3.0.CO;2-Z

*Suh, J., & Moyer, P. (2007). Developing students' representational fluency using virtual and physical algebra balances. *Journal of Computers in Mathematics and Science Teaching, 26*(2), 155–173.

*Taylor, F. M. (2001). *Effectiveness of concrete and computer simulated manipulatives on elementary students learning skills and concepts in experimental probability* (Unpublished doctoral dissertation). University of Florida, Gainesville, FL.

*Threadgill-Sowder, J. A. & Juilfs, P. A. (1980). Manipulative versus symbolic approaches to teaching logical connectives in junior high school: An Aptitude × Treatment interaction study. *Journal for Research in Mathematics Education, 11*(5), 367–374. doi:10.2307/748627

Tindall-Ford, S., & Sweller, J. (2006). Altering the modality of instructions to facilitate imagination: Interactions between the modality and imagination effects. *Instructional Science, 34*(4), 343–365. doi:10.1007/s11251-005-6075-5

Uttal, D. H., O'Doherty, K., Newland, R., Hand, L. L., & DeLoache, J. (2009). Dual representation and the linking of concrete and symbolic representations. *Child Development Perspectives, 3*(3), 156–159. doi:10.1111/j.1750-8606.2009.00097.x

*Wallace, P. (1974). An investigation of the relative effects of teaching a mathematical concept via multisensory models in elementary school mathematics. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 35,* 2998–2999.

*Weber, A. W. (1970). Introducing mathematics to first grade children: Manipulative vs. paper/pencil. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 30,* 3373–3374.

The White House, Office of the Press Secretary. (2009). *President Obama launches "Educate to Innovate" campaign for excellence in science, technology, engineering & math (STEM) education* [Press release]. Retrieved from http://www.whitehouse.gov/the-press-office/president-obama-launches-educate-innovate-campaign-excellence-science-technology-en

*Witzel, B. S., Mercer, C. D., & Miller, D. (2003). Teaching algebra to students with learning difficulties: An investigation of an explicit instruction model. *Learning Disabilities Research & Practice, 18*(2), 121–131. doi:10.1111/1540-5826.00068

*Wood, C. M. (1974). *A comparison of the effects of sequence and mode upon the initial acquisition, retention, and transfer of elementary multiplication concepts.* (ERIC Document Reproduction Service No. ED108909)

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock–Johnson Psycho-Educational Battery–Revised*. Allen, TX: DLM.

*Yuan, Y., Lee, C.-Y., & Wang, C.-H. (2010). A comparison study of polyominoes explorations in a physical and virtual manipulative environment. *Journal of Computer Assisted Learning, 26*(4), 307–316. doi:10.1111/j.1365-2729.2010.00352.x

Received January 4, 2012
Revision received October 16, 2012
Accepted November 1, 2012 ■