A Meta-Analysis and Quality Review of Mathematics Interventions Conducted in Informal Learning Environments with Caregivers and Children

Gena Nelson  
*University of Oregon*

Hannah Carter  
*Boise State University*

Peter Boedeker  
*Baylor College of Medicine*

Emma Knowles  
*Boise State University*

Claire Buckmiller  
*Boise State University*

*See next page for additional authors*

**Publication Information**


This article is protected by copyright and reuse is restricted to non-commercial and no derivative uses. Users may also download and save a local copy for the user's personal reference.

Copyright © 2023, AERA. Published on behalf of AERA by SAGE Publications. Reprinted by permission of SAGE Publications. https://doi.org/10.3102/00346543231156182
For Peer Review
A Meta-Analysis and Quality Review of Mathematics Interventions Conducted in Informal Learning Environments with Caregivers and Children

<table>
<thead>
<tr>
<th>Journal</th>
<th>RER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>Draft</td>
</tr>
<tr>
<td>Manuscript Type</td>
<td>Manuscript</td>
</tr>
</tbody>
</table>

Review of Educational Research

A Meta-Analysis and Quality Review of Mathematics Interventions Conducted in Informal Learning Environments with Caregivers and Children

http://mc.manuscriptcentral.com/rer
A Meta-Analysis and Quality Review of Mathematics Interventions Conducted in Informal Learning Environments with Caregivers and Children
INFORMAL MATH ENVIRONMENTS

Abstract

The purposes of this study were to conduct a meta-analysis and review of study reporting quality of math interventions implemented in informal learning environments (e.g., the home) by children’s caregivers. This study included 23 preschool to third-grade math interventions, and the meta-analysis of 81 effect sizes yielded a statistically significant summary effect ($g = 0.28$, 95% CI [0.08, 0.48]) on children’s math achievement. Significant moderators of the treatment effect included the intensity of caregiver training and type of outcome measure. There were larger average effects for caregiver training that included follow-up support and outcomes that were comprehensive early numeracy measures. Studies met 58.6% of reporting quality indicators and analyses revealed that quality of reporting has improved in recent years. The results of this study offer several recommendations for researchers and practitioners, particularly given the recent, growing, evidence base of math interventions conducted in informal learning environments.

Keywords: informal learning environment, home math environment, intervention, caregiver, math
A Meta-Analysis and Quality Review of Mathematics Interventions Conducted in Informal Learning Environments with Caregivers and Children

Mastering math concepts early has long-term implications for children related to their academic and post-secondary success. Early math development is associated with later math achievement in school (Geary et al., 2018; Kiss et al., 2019; Koponen et al., 2019), as well as various other early-developing skills, such as literacy (Duncan et al., 2007; Purpura et al., 2019). Perhaps more noteworthy, children who encounter difficulties with math early are likely to continue experiencing difficulties in later years (Chu et al., 2019; Morgan et al., 2016). Math deficits can even adversely affect students into adulthood. For example, post-secondary school enrollment and completion rates, as well as employment rates, are related to children’s math proficiency (Davis-Kean et al., 2021; Gaertner et al., 2014; Lee, 2012).

On the 2020 National Assessment of Educational Progress (NAEP), U.S. scores were lower than in 2012 for 9-year-olds (10th, 25th percentiles) and 13-year-olds (10th, 25th, 50th percentiles) (NCES, 2020). The Trends in International Mathematics and Science Study (TIMSS) reported that while U.S. 4th and 8th graders scored higher in math than most participating education systems in 2019, relatively large score gaps existed between the top- and bottom-performing students in both grades (NCES, 2019). Score gaps also increased from most prior TIMSS administrations. These gaps in math knowledge are not limited to school-aged children, and in fact, exist prior to school entry (Burchinal et al., 2011; Chu et al., 2019).

One way to minimize these gaps is extending and enhancing children’s experiences with math outside formal school settings. Developing early math knowledge and skills involves complex processes that begin prior to school entry (Sarama & Clements, 2009). Children, birth to 8 years old, spend more time in informal learning environments (e.g., the home, after school...
INFORMAL MATH ENVIRONMENTS

programs, libraries, everyday experiences such as grocery shopping) than formal learning environments, such as school classrooms (Hofferth & Sandberg, 2001). For all school-aged children, a larger portion of children's time is still spent in informal learning environments compared to the 33 hours per week spent in school (NCES, 2007-2008). This positions adults in caregiver roles (e.g., parents, grandparents, childcare providers) as critical contributors to children’s math development in early years of learning (Ginsburg et al., 2012). Thus, effective caregiver-implemented math interventions rooted in empirically-validated strategies are needed. Throughout this study, we use the term “intervention” to refer to any change in the typical routine of the informal learning environment that caregivers make as a result of receiving information or recommendations in a research study.

Recently, there has been an upward trend in empirical research that investigates the effectiveness of math interventions in informal learning environments (e.g., Berkowitz et al., 2015; Leyva et al., 2018; Niklas et al., 2016; Pupura et al., 2021). Given the long-term benefits of early math learning, a growing need exists to understand the characteristics of effective math interventions in informal learning environments and their potential benefits for young learners. The current meta-analysis included 23 interventions focused on early math content from preschool (e.g., counting, comparison) through 3rd grade (e.g., basic fact fluency, comparing fractions) that were implemented by caregivers in informal learning environments. This is an emerging evidence base, so an essential step to improve research conducted with young children and caregivers is to provide recommendations for improving study reporting quality. Thus, in addition to a meta-analysis, we also investigated the overall quality of the research report (i.e., we examined the degree to which authors reported essential and desirable information in their research reports; referred to throughout as “study reporting quality”).

http://mc.manuscriptcentral.com/rer
Importance of Early Math Development

Early math skills, such as counting and comparison, are building blocks to learning more complex skills, such as addition and subtraction. Children who display difficulty with these foundational skills may continue to have difficulty in other areas of math, such as computation and fraction skills (e.g., Geary et al., 2012). Kiss and colleagues (2019) found that early math skills (e.g., decomposing, numeral identification, number sequence) measured in 1st grade were significant predictors of algebra, data analysis, and geometry skills measured in 3rd grade. Koponen et al. (2019) reported that counting skills measured in kindergarten were predictive of 7th grade overall math performance. Claessens and Engel (2013) found children’s proficiency with identifying numerals, counting beyond 10, and recognizing patterns were the most consistent and significant predictors of later reading and math achievement across elementary school. Similarly, simple retrieval of basic facts and decomposing assessed in 1st grade accurately predicted performance with word problems, whole and rational number computation, and fraction comparison skills in 7th grade (Geary et al., 2013). The results from longitudinal studies suggest that children who do not develop early math skills at school entry are likely to have difficulty with achievement in math in later grades (Nelson & Powell, 2018).

Although many underlying factors impact math achievement, children from low socioeconomic status (SES) backgrounds may have an increased risk of low math achievement. Fewer children who were eligible for free or reduced-price lunch (FRL) scored Proficient or Advanced (26%) in math compared to children who were not eligible for FRL (58%) on the 2019 NAEP (NCES, 2019). Morgan et al. (2009) found that children from families of lower SES had slower growth in math between kindergarten and 5th grade compared to children from higher SES families. Students identified as English Learners (ELs) may also be at increased risk of low
math performance. NAEP data show an achievement gap between ELs and non-ELs in math over the last two decades. Only 16% of 4th grade ELs scored Proficient, while 44% of non-ELs scored Proficient (NCES, 2019). Similarly, 4th and 8th grade students with disabilities have lagged behind peers in math for decades (NCES, 2019). Because gaps in math achievement have persisted, researchers are exploring factors that may decrease these gaps. For example, Galindo and Sonnenschein (2015) found that supportive home math environments (HME) can significantly decrease achievement differences related to SES. While many factors likely contribute to achievement gaps, children considered at-risk academically may especially benefit from math interventions conducted in informal learning environments.

The Role of Informal Learning Environments

Informal learning offers children opportunities related to “how, what, where, and with whom they learn” (Pattison et al., 2017, pp. 6), regardless of the environment. In other words, informal learning can take place anywhere. Informal learning environments, on the other hand, are spaces such as the home, out of school programs, the outdoors, sports, and actions such as exploring exhibits at libraries and museums, and engaging in everyday experiences (e.g., cooking) (Gerber et al., 2001). This study prioritizes children’s learning in informal learning environments, and we consider both studies that focus on facilitated or designed learning (e.g., math-themed museum exhibits) and unstructured learning (e.g., free play with blocks).

Recent initiatives recognize the importance of informal learning environments. For example, the Mathematics Learning in Early Childhood: Paths Toward Excellence and Equity report (NRC, 2009) recommended education partnerships should form between families and community programs to promote math learning. In 2002, the National Association for the Education of Young Children and the National Council for Teachers of Mathematics released a
joint position statement with a similar emphasis. This recommendation is not surprising, since young children come to school with some knowledge of letters and numbers, likely resulting from caregivers providing children with learning opportunities in informal learning environments (Passolunghi & Lanfranchi, 2012; Susperreguy & Davis-Kean, 2016). In a meta-analysis of 64 studies focused on the relation between the HME and children’s math achievement, authors reported a statistically significant relationship ($r = .13, p < .001$; Daucourt et al., 2021).

Similarly, Mutaf-Yildiz et al. (2020) reported significant relations between the HME and math achievement in 33 of 37 studies. The positive relation between the HME and children’s math achievement, and the amount of time children spend in informal learning environments highlights the importance of identifying effective math interventions caregivers can implement.

**Math Learning in Informal Learning Environments**

The body of research examining the role of informal learning environments and student achievement is growing in the U.S and worldwide (e.g., Manolitsis et al., 2013; Napoli & Purpura, 2018; Sheldon & Epstein, 2005). LeFevre et al. (2009) found that HME activities from kindergarten to 2nd grade accounted for a significant proportion of variance in math knowledge and math fluency skills, even after controlling for other variables including vocabulary, working memory, and the home literacy environment. When examining the relationship between home literacy and math practices and children’s academic achievement, the HME was predictive of numeracy and vocabulary outcomes (Napoli & Purpura, 2018). In a longitudinal study, Manolitsis et al. (2013) examined the role of home literacy and numeracy experiences on reading and math outcomes for children in kindergarten and 1st grade; parent literacy teaching (e.g., letter names) predicted early math achievement as strongly as parent numeracy teaching.

Sheldon and Epstein (2005) examined eight family and community involvement activities across
INFORMAL MATH ENVIRONMENTS

elementary and high school (e.g., parent workshops on math skills, parent conferences to discuss children’s progress in math). Only one type of involvement (home learning activities) consistently and significantly related to children’s higher math achievement.

Caregivers in informal learning environments, then, play a critical role in children’s math development (Ginsburg et al., 2012). Interventions in informal learning environments are often led by adults who have no prior experience with teaching or math, making treatment fidelity and interventionist training particularly important. Accurate interpretation of treatment effects and recommendations related to generalization depend on adherence to the quantity and quality of intervention activities (Capin et al., 2018). A vast body of research shows that when given the opportunity to play board games (Laski & Siegler, 2007; Siegler & Ramani, 2009), read books focused on number concepts (Hendrix et al., 2019; Purpura et al., 2017), and engage with math education apps (Griffith et al., 2019; Zippert et al., 2019), children in both informal and formal learning environments increase their math knowledge. This is particularly true in the preschool years, which is encouraging since early math learning promotes greater achievement more broadly, such as in literacy and science (Duncan et al., 2007).

Previous Reviews of Math Intervention Studies

Researchers have conducted syntheses to identify effective components of math interventions in classroom settings at various grade levels. Nelson and McMaster (2019b) focused their meta-analysis on 34 studies of early numeracy interventions for children in preschool through 1st grade, reporting a large (Kraft et al., 2020) summary effect ($g = 0.64$). Similarly, Charitaki et al. (2020) evaluated numeracy interventions for preschool to 2nd grade students and found a similarly large effect size ($g = 0.61$); the authors reported larger treatment effects for short-term interventions including 1 to 9 sessions. Related to students with
exceptional learning needs, Kroesbergen and Van Luit (2003) investigated 58 math intervention studies for elementary children with low math achievement; authors reported a large summary effect ($d = .92$) for studies that focused on early math skills (i.e., counting, number sense).

Gersten et al. (2009) reviewed math intervention studies for kindergarten through 12th grade children with learning disabilities and reported a large effect ($g = .63$). Although the results of these syntheses support the use of interventions to improve math performance, only studies conducted in classrooms with teachers or researchers as interventionists, rather than in informal learning environments with caregivers delivering the intervention, were examined.

Researchers have conducted few syntheses of interventions with caregivers acting as interventionists (e.g., Fishel & Ramirez, 2005; Van Steensel et al., 2011). Van Steensel et al. (2011) conducted a meta-analysis of 30 studies that examined the effectiveness of home literacy programs with preschool and primary school children. Results indicated a small but statistically significant mean effect ($d = 0.18$). Fishel and Ramirez (2005) reviewed 24 studies that examined the effectiveness of parent tutoring on kindergarten through 12th grade children’s academic performance, including reading, math, spelling, and homework completion. Although this systematic review included math interventions, the authors did not include participants in the preschool age-band or conduct a meta-analysis, and their literature search ceased in 2003. These analyses offer meaningful insights for young learners; however, to our knowledge, there is not a meta-analysis that specifically aims to investigate the effects of early math interventions implemented in informal learning environments.

**Study Reporting Quality**

In addition to conducting a meta-analysis, we also reviewed study reporting quality. One criticism of meta-analysis is “garbage in, garbage out,” meaning if meta-analyses include several
low quality studies, flaws of the primary studies can be carried over into the meta-analysis (Borenstein et al., 2009). Thus, many meta-analysts have coded for study reporting quality and considered it in their analysis (e.g., Jitendra et al., 2021). Although the quality of study reporting is not necessarily a precise measure of the quality of a study’s findings (i.e., if it can be considered “garbage”), studies with high quality of reporting can reveal limitations to the generalizability of findings and low quality of reporting may raise concerns regarding the validity of intervention effects. The evidence-base for math interventions in informal learning environments is still emerging, so now is also an appropriate time to conduct a comprehensive evaluation of study reporting quality. With any evidence base, researchers and practitioners benefit from understanding the current limitations of the field. Evaluating study reporting quality can support researchers in identifying current limitations of how information about interventions (e.g., content, implementation) and the effectiveness of interventions is reported. The results of a quality review provide essential recommendations to enhance study reporting, thereby allowing researchers and practitioners to determine if interventions are representative, generalizable, and transferrable to the context in which they support children and families (Cook & Cook, 2017).

Previous researchers have examined study quality in math intervention meta-analyses. The studies reported varying results, with average study reporting quality ratings (scale of 0 to 1) ranging from, for example, 0.77 for geometry interventions for 4th to 12th grade students with learning disabilities (Liu et al., 2019), to 0.83 for tier 2 math interventions for students with math difficulty (Jitendra et al., 2021), to 0.87 for proportional reasoning interventions for students with math difficulty or learning disabilities (Author et al.). Wide ranges in study reporting quality across math interventions within the same meta-analysis are reported (e.g., Stevens et al., 2018). To our knowledge, researchers have not conducted a review of study reporting quality of math
interventions conducted in informal learning environments.

**Purpose and Research Questions**

This project addresses an important gap in the literature—features of effective methods that caregivers use to promote young children’s math understanding. The present study contributes (1) a meta-analysis of research on the effectiveness of early math interventions in informal learning environments, and (2) an evaluation of the current state of study reporting quality of math interventions in informal learning environments. The research questions were:

1. What is the average treatment effect, and how variable are the effects of math interventions implemented in informal learning environments?
2. What variables (e.g., intervention characteristics, outcome measure content, intervention activities) predict effectiveness for the total sample of studies?
3. What is the reporting quality for studies included in the meta-analysis, and what trends exist across quality indicator variables?

**Method**

The Method section provides a detailed overview of the following aspects of this meta-analysis: (a) description of the inclusion criteria, (b) procedures we used to conduct the literature search, (c) training procedures for the abstract screening and full-text review, (d) the results of the literature search procedures, (d) article coding procedures, (e) the training for and reliability of the coding procedures, and (f) the data analysis plan.

**Eligibility for Inclusion in the Meta-Analysis**

We applied the following a priori inclusion and exclusion criteria throughout the literature search process to identify studies for the meta-analysis. Studies:

1. Investigated math intervention effects. We defined intervention as a change to the typical
routine such as researcher recommendations to focus on specific math knowledge and skills in the home or engage in an activity for a specified amount of time. We included intervention programs that primarily focused on math, as opposed to early childhood or elementary holistic curricula programs (Noble et al., 2012).

2. Were conducted in informal learning environments, defined as home, out-of-school programs held at community spaces such as libraries and museums, and everyday experiences (Gerber et al., 2001). We excluded studies conducted only in formal learning settings, including classroom interventions and studies at university research labs or lab schools (Brown & Alibali, 2018). However, if a study included both formal and informal learning environment components, the informal learning environment component was included if the effects of that intervention could be isolated (Starkey et al., 2004).

3. Administered at least one math achievement outcome measure to determine intervention effectiveness. Studies that only administered measures of “math talk” were excluded (Vandermaas-Peeler et al., 2016), as were studies that only administered cognitive measures that are not math achievement measures (e.g., executive functioning, spontaneous focus on numerosity (SFON); Braham et al., 2018). Math talk measures were not considered for this meta-analysis because the effect sizes derived from this outcome are different from typical math achievement outcomes, and therefore, cannot be aggregated with achievement outcomes.

4. Had an adult in a caregiver role (e.g., parent, grandparent, older sibling, childcare staff at a library or museum) as the primary facilitator of the child-directed components of the math intervention. Researchers could serve as the implementer of the parent-directed component of the intervention (i.e., leading the caregiver workshop on math activities
that caregivers then implemented at home with their child).

5. Included child participants between 3 years, 0 months (average age at the start of the intervention) to the end of 3rd grade (average age less than 9 years, 0 months old). Studies with younger or older children were excluded (Lynch & Kim, 2017) unless data were disaggregated for children within the target age.

6. Used an experimental or quasi-experimental group design. Studies that used single case design or qualitative methods were excluded (Linder & Emerson, 2019).

7. Provided appropriate information to calculate effect sizes (e.g., means, SDs, F statistics, t-tests). Studies that did not report information to calculate effect sizes (or could not be obtained directly from authors) were excluded (e.g., Pasnik et al., 2015).

8. Presented results in English.

Literature Search Steps

We conducted a comprehensive literature search process to identify studies for inclusion in the meta-analysis. We reviewed peer-reviewed journal articles, dissertations, conference presentations, book chapters, and technical reports, with no restrictions on publication year or type. We captured relevant studies with (a) electronic databases, (b) journal table of contents, (c) direct email communication with experts, (d) emails to organizational email listservs, (e) preprints, (f) reference lists, and (h) forward citation search.

Electronic Databases Search

First, we searched electronic databases including: Academic Search Premier, Education Research Complete, ERIC, ProQuest Dissertations & Theses, PsycARTICLES, and PsycINFO with the following Boolean search string: (intervention OR activity OR training OR tutoring OR “book reading” OR tablet OR “e-book” OR “math* play” OR “informal learning”) AND (math*
INFORMAL MATH ENVIRONMENTS

OR numeracy OR "number sense" OR “math talk” OR geometry OR algebra) AND (parent* OR childcare OR caregiver OR daycare OR "day care" OR day-care OR "after school" OR museum OR "home tutoring" OR home-based OR "home learning environment" OR "home math*

environment" OR “home numeracy”) AND (preschool* OR prekindergarten OR "early childhood" OR kindergarten OR "first grade" OR "second grade" OR "third grade" OR elementary OR “primary school” OR “head start” OR “nursery school”).

Journal Table of Contents

Second, we searched the electronic table of contents for relevant journals (using the same Boolean search string as the electronic databases) for the following 10 journals: Curator: The Museum Journal; Early Child Development and Care; Early Education and Development; Journal of Cognition and Development; Journal of Research in Childhood Education; Journal of Research on Educational Effectiveness; International Journal of Early Years Education; International Journal of Science Education; Mind, Brain, and Education; and Visitor Studies.

We also conducted a hand search of journals (publication years 2016 to 2021) that did not allow for an electronic search using the Boolean search string, including five journals: Children and Libraries; Cognitive Development; Early Childhood Research Quarterly; Journal of Experimental Child Psychology; and Journal of Research in Mathematics Education.

Email Communication with Experts in the Field

Third, we emailed 20 researchers who regularly conduct research related to the topic of this meta-analysis. We provided the researchers with the inclusion criteria and asked them to send any published or unpublished studies from their research labs that met inclusion criteria.


**Researcher Contacts Through Relevant Organizations**

Fourth, we contacted organizations maintaining an email listserv with subscribers who may have conducted studies aligned to the focus of this meta-analysis. We contacted several organizations, and the following posted our announcement to their listserv: Association of Science-Technology Center, Cognitive Development Society, International Group for the Psychology of Mathematics Education, Mathematics Cognition and Learning Society, and Visitor Studies Association. We also emailed the National Science Foundation’s Center for the Advancement of Informal Science Learning for information on former grant projects.

**Pre-Print Publications**

Fifth, we searched PsyArXiv (https://psyarxiv.com/) for pre-prints of studies that authors may have submitted to the repository. Based on the electronic search function of this database, we searched two-word combinations of the following terms: math, numeracy, intervention, tutoring, informal, activity, play, and home.

**Reference List Search of Included Studies**

Sixth, we searched the reference lists of three included studies that met inclusion criteria and because they were under review at journals, they represented the most recent research. They also represented different types of interventions, including text messaging (Napoli & Purpura, 2021), picture books (Purpura et al., 2021), and food routines (Leyva et al., 2021).

**Forward Citation Search**

Finally, we conducted a forward citation search of the three most cited publications, according to Google Scholar’s “cited by” function, out of all the included studies in this meta-analysis (Niklas et al., 2016; Starkey & Klein, 2000; Vandermaas-Peeler et al., 2012).
Literature Search and Review Procedures

The abstract screening and full-text review began in January 2021, ended in August 2021, and were organized and supervised by the first and second authors. Three graduate assistants (GAs) conducted the literature search. We led four 1.5 hour introductory training sessions for the GAs on the purpose and defining features of the meta-analysis, features of published early math intervention studies, and best practices for abstract screening. The GAs were pursuing master’s degrees in Counseling, Early and Special Education, and Elementary Education. One GA had previous experience conducting literature reviews.

Training

Abstract Screening

We held a 2-hour training session that included an overview of the abstract review criteria for this meta-analysis, the process for filling out the Excel database to identify which studies should be reviewed in full-text, and a group practice of applying the criteria to abstracts. Through an iterative process of applying criteria to 175 practice articles and discussing challenges with the screening tool, we finalized the screening tool. After training and practice, the GAs independently coded the remaining abstracts. All abstracts were double-coded.

Throughout the abstract screening process, we held weekly 1-hour meetings to discuss disagreements about inclusion, increase screening process reliability, and provide continuous GA follow-up support. Each week, GAs coded a set of 400-600 abstracts, and the third author calculated Kappa. Kappa is a stricter measure of interrater reliability than percent agreement and takes into account the possibility of rater guessing. Values of Kappa between 0.60 and 0.79 are considered “moderate” (McHugh, 2012). Kappa improved from the first two sets of abstracts (0.25 and 0.42, respectively) to the final two sets of abstracts (0.78 and 0.71, respectively). All
discrepancies were discussed during weekly meetings resulting in 100% agreement on the final set of included and excluded studies.

**Full Text Review**

A similar training process as the abstract screening was used for the full-text review (initial training, iterative feedback loop, weekly meetings for follow-up support). The full-text review included GAs reading and scanning sections of the full-text of articles and capturing data from the studies (e.g., children’s age, intervention setting, outcome measures) so the first and second author could make final determinations about the inclusion of studies. Two GAs independently coded all studies for full-text criteria. Each GA coded 30-40 full text articles each week. Discrepancies not discussed as part of the research team meetings were resolved between the GAs in individual meetings. Interrater agreement was 95.3% for full-text coding.

**Literature Search Results**

Figure 1 is a PRISMA diagram that documents the literature search results, including the abstract screening and full-text review results. Total, we captured and reviewed 7,245 abstracts. At the abstract screening stage, most studies were excluded because they did not: focus on academic learning (or were generally irrelevant to early learning), represent a research study, include an intervention, have independent variables related to mathematics, use a group design, include participants in PK-3rd grade. Duplicated studies were also removed.

We identified 373 studies for full text review. We excluded studies because they did not: take place in informal learning environments; focus on testing the effectiveness of an intervention; provide child participants with mathematics content intervention, play, or support; use a group design method; include child participants between the ages of 3 and 9 years; administer a mathematics achievement measure; present effects of the home component that
could be isolated from the effects of the school intervention; provide enough information to appropriately calculate effect sizes. Duplicated studies were also removed (e.g., conference presentations that had eventually been published in a peer-reviewed journal).

As we reviewed studies in the full-text phase, the first and third authors made notes about methodological concerns for studies. We enforced ad hoc exclusion criteria to ensure that the studies included in this meta-analysis were of high methodological quality. We excluded one study with severe levels of attrition (i.e., more than 50%; Lore et al., 2016), one study that employed a treatment group only design (Leyva et al., 2018), and one study that used nonrandomization procedures without administering a pre-test and also lacked specific information needed to estimate the effect of clustering on variance estimates (Austin, 1988). We excluded two studies (Flynn, 2021; Libertus et al., 2020) because the math achievement measures the authors administered did not specifically align with the math content of the intervention. This particular situation is problematic because the goal of the current meta-analysis was to synthesize the effects of interventions on the outcomes that the interventions were intended to effect. For example, Flynn (2021) included a comprehensive measure of early numeracy, operations, measurement, contextual problems, spatial awareness, etc., but the intervention was a narrowly-focused counting game. The outcome measure in this study is not aligned with the content of the intervention; therefore, the effect size would not accurately represent the effect of the intervention on the intended math content. Whereas, other studies that used comprehensive measures did so because the interventions addressed multiple math skills. In total, we included 23 intervention studies that represented appropriately rigorous methodology.

**Coding Procedures**

The first, second, and third authors drafted the coding protocol; authors have extensive
experience creating coding protocols for systematic reviews, meta-analyses, and qualitative studies. We used an iterative process of feedback from GAs, advisory board members, and consultants to refine and finalize the coding protocol. The coding protocol comprises five sections (basic study information and participant demographics, methodological information, intervention features, outcome measures, effect sizes) and is briefly explained in Table 1. The full coding protocol contained nearly 200 variables and is on an open access platform (link blinded for review). The coding protocol included a variable name, code options (e.g., forced response, open response), and the variable description or definition, with examples as applicable.

**Quality of Study Reporting**

Within each of the five sections of the coding protocol, we embedded study reporting quality codes, referred to as quality indicators (QIs). The coding protocol for QIs is also briefly explained in Table 1, with each individual indicator reported in the results. We coded studies for 37 QIs based on Gersten et al. (2005), which provided recommendations for reporting standards in experimental and quasi-experimental studies. We coded for QIs related to *reporting* quality, meaning that we recorded whether studies provided specific information to support readers’ replicability and understanding of the method and interpretation of the results (as we had previously removed studies due to methodological issues). Gersten et al. (2005) emphasized components related to special education research, but the QIs are highly relevant for all educational intervention studies. We also referred to PRISMA reporting guidelines (Page et al., 2021) to add relevant variables related to study reporting quality to the coding protocol.

The 37 QIs included variables that Gersten et al. (2005) deemed (a) essential \(n = 27\) and (b) desirable \(n = 10\). We used the essential and desirable framework, as essential QIs may be necessary for researchers and practitioners to replicate and generalize the intervention study,
whereas the desirable QIs represent supplemental information that provide helpful study detail.

We coded QIs across the following main categories: 1.0 participant description, 2.0 intervention and comparison condition descriptions, 3.0 outcome measures, and 4.0 data analysis and results.

**Training and Reliability for Coding**

The second author led coder training for the three GAs responsible for coding all included studies for sections 1-4 (including embedded QI codes). The first and third author completed the section 5 coding (extracting effect size information). The initial training included reviewing the variables and details of the section of the coding protocol, practicing with a small set of studies, and discussing challenges with the codes. All studies used for practice were coded by all GAs, and the remaining studies were each coded by two GAs. At the end of training, interrater reliability using percent agreement for the sections of coding was 0.82 (section 1: study-level variables), 0.97 (section 1: child-level variables), 0.86 (section 2: methodological information), 0.89 (section 3: intervention features), 0.82 (section 4: outcome measures), and 1.00 (section 5: effect sizes). The study quality codes were embedded within the five sections.

**Data Analysis Plan**

*Meta-Analysis Procedure*

We calculated Hedge’s $g$ effect sizes for each study using procedures outlined in the What Works Clearinghouse statistical guidance documentation (WWC, 2017) and found in meta-analysis support materials (Lipsey & Wilson, 2001; Morris & DeShon, 2002). This required calculation of Cohen’s $d$ based on means, standard deviations, and sample sizes and then correction for small sample size (thus producing Hedge’s $g$; Hedges & Olkin, 1985). All effect sizes were derived such that a positive value indicates the intervention group scored on average higher than the control group. When pre/post tests were used to calculate an effect size, the
correlation between the pre-test and post-test was required for effect size variance estimation. If this was not presented in the article, the authors were contacted. If authors were unresponsive, a value of 0.5 was used in accordance with WWC guidelines (WWC, 2017).

We initially identified potential outliers in effect sizes as those that were more than +/- 3.0 SDs of the random-effects weighted mean effect size (Cooper et al., 2009). We used a funnel plot and Egger’s test (Egger et al., 1997) to assess for potential publication bias. Egger’s test requires specifying a meta-analysis model in which the effect sizes are predicted by their standard errors. Finally, we used a contour-funnel plot to flag studies that may be outliers due to factors other than publication bias.

We used random-effects models over fixed effect models based on the assumption that a distribution of effects exists in the population (Borenstein et al., 2009). Several studies reported more than one effect size (i.e., studies included more than one math achievement measure), creating a dependency between effect sizes that must be accounted for to produce accurate statistical tests. To account for this type of nested data structure and dependence, we conducted multilevel meta-analyses with robust variance estimation (RVE; Hedges et al., 2010; Pustejovsky & Tipton, 2021) using the R packages *metafor* (Viechtbauer, 2010) and *clubSandwich* (Pustejovsky, 2019). RVE adjusts standard errors to account for correlations between effects within studies and is possible without precise knowledge of the correlations between effects; a value of 0.80 is assumed and sensitivity to this value assessed (Hedges et al., 2010).

Variability in the distribution of effects supports investigation of potential moderators. We estimated heterogeneity ($\tau^2$) using restricted maximum likelihood (Boedeker & Henson, 2020) and variability in potential future effects is described with the 95% prediction interval. Whereas the confidence interval for the average effect provides an indication of its precision and
statistical significance, the prediction interval provides the expected range of true effects in similar studies conducted in the future (IntHout et al., 2016). The $Q$ test is used to assess the statistical significance of the variability across effect sizes.

To explain heterogeneity, we tested several moderators. Moderators included: (a) content of the math achievement outcome (number identification, addition, number line estimation, quantity discrimination, missing number or number order, comprehensive early math ability (e.g., TEMA), comprehensive early numeracy (multiple numeracy skills), and non-numerical skills); (b) intervention characteristics (amount of caregiver training, intervention location, number of intervention sessions, length of intervention); and (c) activity (children’s books, number board games, card games, dice games, other counting or number games with varying options for activities, food experiences, formal caregiver curriculum). Individual moderators were assessed using the RVE corrected multilevel meta-analysis model. Not all studies reported all moderators. Therefore, multiple imputation was utilized to impute missing moderator values using mice (van Buuren & Groothuis-Oudshoorn, 2011), and results were compared to complete case analyses. Results were similar; therefore reported results are from complete case analyses.

**Study Reporting Quality**

Each QI was scored on a scale of 0 (not met) to 1 (met), similar to previous reviews of quality using these indicators (e.g., Jitendra et al., 2021; Stevens et al., 2018). We calculated average scores for each QI variable, QI category, and study. Within the text of the results section we refer to the average scores as percent of QIs met to aid in interpretability of the scores.

**Results**

**Descriptive Results**

The 23 studies included 21 publications (two publications included two studies each) and
were published from 1981 to 2022, with some studies under review as of February 2022. At the time of article coding, most studies were published peer-reviewed journal articles (74%; \( k = 17 \)), with three studies (13.0%) in progress or under review at peer-reviewed journals. Three studies (13.0%) were masters’ or doctoral theses. The majority of studies (74%; \( k = 17 \)) were conducted in the U.S., and all studies were implemented solely or partially in the home learning environment. In all studies and intervention groups, the child’s primary caregiver was the interventionist. Supplementary table S1, in the online version of this journal, provides an overview of the 23 studies, including the math content and intervention activities.

There were 3,292 total child participants with an average age of 4.78 years, ranging from 1.58 years to 8.5 years. Supplementary Table S2 presents a summary of the participants across the 23 studies, including the number of studies that reported specific demographic information. For example, 16 of the 23 studies reported information about participants’ race or ethnicity. Using race and ethnicity categories that are commonly applied to research studies conducted in the U.S., the table shows that 9 studies included participants who were Black, for a total of 304 child participants (which is equivalent to 44.77% of participants from U.S. studies). We did not identify any studies that reported including children with disabilities, and we identified only one study that reported including children who were dual language learners.

**Meta-Analysis Results**

**Overall Effect**

For quantitative analysis, we included 81 effect sizes from 23 studies (average of 3.5 effects per study). The overall effect was statistically significant, \( g = 0.28, 95\% \text{ CI } [0.08, 0.48], t(17.9) = 2.96, p = 0.010 \). The fractional degrees of freedom is due to the application of RVE with a small sample correction (Tipton, 2015). To provide some context to this finding,
educational interventions with effect sizes larger than 0.20 may be considered “large” (Kraft, 2020). A caterpillar plot showing the distribution of effect sizes can be found in Figure S1.

**Variability in Effects**

The variance of the distribution of effects was statistically significant, $\tau^2 = 0.068$, $Q(80) = 104.27$, $p = 0.04$. With a variance estimate of 0.068, the estimated standard deviation of the distribution of effects is 0.26 and the 95% prediction interval spans a wide range of values from -0.25 to 0.81. Given the magnitude of the variability in effects and the potential for future studies to produce either positive or negative effects, understanding the intervention characteristics and activities most likely to lead to positive and substantial gains is warranted.

**Publication Bias**

The funnel plot did not show substantial asymmetry (see Figure 2), though values outside the funnel were further scrutinized and ultimately retained. Egger’s test was implemented to evaluate funnel plot asymmetry. For consistency with the other models, a multilevel model with RVE was specified for Egger’s test. The result was a non-statistically significant ($p = 0.448$) coefficient of the standard error of the effect sizes; therefore, we fail to reject the null hypothesis of a zero relationship between the standard error and effect size. A contour-enhanced funnel plot, useful for detecting publication bias resulting from the suppression of statistically insignificant findings, does not indicate substantial issues (see Figure 2).

**Moderators**

As an initial step, moderators were evaluated individually. Results for categorical moderators required specifying a model with no intercept whereas continuous moderators included an intercept. As evident in Table 2, none of the potential moderators yielded statistically significant differences across their levels and the residual variability remained statistically
significant for all but one of the moderators (intensity of caregiver training). For intensity of caregiver training, \( \tau^2 \) was 0.030, a reduction of 52% in the unexplained variability across effect sizes.

The meta-regression results for each individual moderator are presented in Table 3. All analyses were conducted with and without controlling for quality of reporting variables. Results across the two sets of analyses were consistent. Because quality of reporting refers only to the information that is reported, and may not capture a complete illustration of the research study that authors conducted, we report the results without controlling for quality of reporting. For math outcome measure content, comprehensive early numeracy measures (i.e., a measure that addressed multiple early numeracy skills) was the only category of outcome measures that was statistically significantly greater than zero (\( g = 0.39, 95\% \text{ CI} [0.08, 0.70], p = 0.01 \)). Of the four intervention characteristics evaluated, intensity of caregiver training and principal intervention environment each had a level that was statistically significant. For caregiver training, the use of an initial training session with follow-up support (e.g., additional training sessions, in-home coaching, text message support) was associated with an average effect that was statistically significantly different from zero (\( g = 0.43, 95\% \text{ CI} [0.15, 0.70], p = 0.01 \)), whereas when no training or only initial training occurred, the average effects were non-significant.

The majority of studies examined the effect of caregivers implementing math intervention treatment conditions (\( n = 19 \)) in which the home was used as the sole, primary learning environment. A subset of studies examined the effect of considering two learning environments for treatment conditions (\( n = 5 \)), which included classroom-based interventions with home learning environment components. The purpose of this meta-analysis was to determine the effect of learning that occurs in informal learning environments, therefore, the
main effect of treatment conditions in which children were exposed to two principal learning environments represents only the isolated effect of the informal learning environment (i.e., effect sizes were calculated as classroom intervention vs. classroom intervention plus home component). The average effect was statistically significantly different from zero when the intervention occurred in the home only ($g = 0.31, 95\% \text{ CI } [0.09, 0.54], p = 0.01$). When the informal learning environment serves as a component of an intervention that primarily takes place in a formal learning environment (e.g., the classroom), the additive effect of the informal learning environment is positive but not statistically significant ($g = 0.16, p = 0.18$).

Considering the main activities that caregivers used to implement the intervention with children, the use of children’s math story books was the only activity that yielded a $p$-value less than 0.05; however, given the degrees of freedom for this activity were less than 4 the result should not be characterized as statistically significant. All other activities except varying counting and number activities (i.e., caregivers were given several counting and number activities and games to choose from throughout the duration of the intervention) had average effect sizes larger than children’s books. Regarding math content of the intervention, all average effect sizes were positive but only effect sizes pertaining to interventions that included relations skills ($g = 0.26$) or operations skills ($g = 0.32$) were statistically significant.

**Study Reporting Quality Results**

Table 4 provides the average QI score (percentage of QIs met) for each of the four main QI categories and their respective subcategories and individual QIs. The results for each study can be found in the online supplementary materials, Table S3). Studies met 58.6% ($SD = 13.9\%$) of all QIs. The 23 studies met an average of 70.1% of the essential indicators (range = 38.5% to 88.5% QIs met across studies). On fewer occasions, studies met the requirements for reporting of
the desirable QIs (31.2%). Studies published most recently in the last 3 years (since 2019) had a
higher percentage of essential (76.9%) and desirable (41.3%) QIs met compared to studies
published in 2018 or early (64.9% and 22.0%, respectively).

On average, the 23 studies met 64.9% of the QIs related to participant description (range
of 25.0% to 91.7% across studies), 60.3% of the QIs related to the description of the intervention
and comparison conditions (range of 16.7% to 100% across studies), 31.0% of the QIs related to
outcome measures (range of 0% to 75.0% across studies), and 83.5% of the QIs related to the
data analysis plan and results (range of 60.0% to 100% across studies).

Discussion

The purpose of this study was to conduct a meta-analysis of early math interventions
implemented by caregivers with young children in informal learning environments, with a
subsequent review of study reporting quality of these interventions. The results of the current
meta-analysis may support training and future research focused on caregivers and other early
childhood professionals who support young children’s math development in informal learning
environments. The results of the current quality review reveal strengths and areas of growth for
reporting the results of interventions conducted in informal learning environments.

Implications of the Meta-Analysis Results

The results of the meta-analysis revealed a statistically significant average effect ($g = 0.28, p = 0.010$) of early math interventions conducted in informal learning environments, and an
effect size of 0.28 is considered “large” (Kraft, 2020). To date, there is not another published
meta-analysis on math interventions conducted in informal learning environments; however, we
are aware of another team of researchers with a similar study that is in progress. The authors
reviewed home-based math interventions for 3-to-5-year old children and reported an effect of $d$
= 0.18 (95% CI [– 1.62, +1.99]; n = 8) for math outcomes (A. Cahoon, personal communication, January 13, 2022).

An effect size of 0.28 is impressive for several reasons. The results of this meta-analysis reported a smaller average effect compared to meta-analyses that examined early math interventions conducted in classroom settings (e.g., $g = 0.64$, Nelson & McMaster, 2019b; 0.61, Charitaki et al., 2020). However, many interventions included in previous meta-analyses have been conducted as part of research studies in which researchers or graduate students implement interventions and fidelity of implementation is strictly monitored, potentially leading to larger effects. The interventions included in this meta-analysis were implemented in authentic settings, outside of laboratory settings or the structure of a classroom intervention. Moreover, an average effect of 0.28 is equivalent to an improvement index of approximately 11 percentile points. The improvement index is the expected change in terms of an increase or decrease in percentile rank for an average student in the control group had that student participated in an intervention (WWC, 2017); therefore, our findings indicate that a student at the average of the control group distribution would be expected to increase their percentile rank by 11 points had the student participated in an intervention. These results suggest that caregivers can leverage the home learning environment without significant disruption or adaptation to the typical routine of the home and support children’s math development, and potentially even support the narrowing of math achievement gaps.

The interventions included in this meta-analysis also required very few resources or materials. Caregivers were provided with materials that are typically available for low cost (e.g., dice, number board games), and the results of the current meta-analysis did not suggest that caregivers or children require access to more expensive materials such as iPads to reap the same
benefits of the intervention. The included interventions also required very little time to implement. The average intervention length was approximately 8 weeks, and the recommended intensity ranged from once per week to daily. It is empowering, for researchers and caregivers alike, to observe the positive impact that caregivers can have on children’s math development and understanding with non-intensive interventions. It is encouraging that the results of this meta-analysis suggest that low-cost, easy to implement interventions by caregivers with varying SES backgrounds and different levels of education are effective in increasing young children’s math achievement. The results of this meta-analysis make significant contributions to the literature on methods for reducing math achievement gaps.

**Effect of Caregiver Training**

The majority of interventions in this meta-analysis were non-intensive in their training requirements, which also speaks to the practical significance of the overall effect of $g = 0.28$. The current meta-analysis results show that caregivers can positively impact children’s math development and understanding, even with a minimal time commitment, and this is even more so true when they are trained and supported. Based on our analyses, the use of initial training with follow-up support was statistically significant, whereas when no training or initial training with no follow-up support occurred, the average effects were non-significant. Examples of caregiver training from included studies involved a 30-min presentation at an early childhood center parent night (Niklas et al., 2016), a demonstration viewing to learn how to play a game (Ramani & Scalise, 2020), and three at-home visits for the parent to practice activities with coaching support (Dulay et al., 2019). Examples of follow-up support included text message reminders three times per week (Leyva et al., 2018), researcher phone calls once per week (Libertus et al., 2020), and three classes involving a lesson and time for parent questions (Blanch, 2002).
Interventions with caregiver training and follow-up support can yield greater effects (i.e., be more influential to child learning). This is a promising finding for several reasons. First, training caregivers to implement non-intensive math interventions can be particularly meaningful for children who struggle and need additional practice opportunities to develop proficiency in math. In other words, ensuring that caregivers are provided training and support for learning in the home can contribute to minimizing achievement gaps in math that are prevalent before school entry (Burchinal et al., 2011) and can significantly decrease achievement differences related to SES (Galindo & Sonnenschein, 2015). Second, the interventions included in the current meta-analysis typically required caregivers to attend an initial training session lasting between 30 min and 2 hours. This means that an extensive time commitment is not required by caregivers to have an impact on children’s math achievement. Finally, training can easily be controlled or manipulated as part of interventions in informal learning environments or as part of family engagement in community and school programs. For example, early childhood teachers or community leaders who offer caregiver training can do so with confidence that their efforts will likely have positive impacts on student learning. Because of the impacts on child learning, future intervention research can prioritize enhancing caregiver training and support opportunities, which can in turn increase fidelity of intervention activities. Additionally, meta-analysts may want to pursue future research focused on caregivers and other early childhood professionals who support young children’s math development in informal learning environments.

**Effect of Intervention Outcomes**

The results of the individual moderator analysis reported that comprehensive early numeracy outcomes yielded the only statistically significant effect. Although not to a statistically significant level, comprehensive early math measures (those measures that included more than
just multiple numeracy skills) and measures of non-numerical skills also yielded similarly large
effects (Kraft, 2020). This was a surprising result because previous researchers have reported
that comprehensive measures may not be sensitive enough to detect growth during an
intervention (Nelson & McMaster, 2019b). The positive impacts on math achievement broadly
give confidence in the ability of non-intensive math interventions implemented in authentic
settings to contribute to the overall growth of children’s math knowledge and skills.

**Effect of Primary Intervention Location**

In order to be included in this meta-analysis, any intervention study with a formal
learning environment component (e.g., the classroom) and an informal component (e.g., the
home) had to provide results in a manner that allowed us to isolate the effects of the intervention
specific to the informal learning environment. It is important to acknowledge that although the
interventions conducted in only the home environment yielded a larger effect ($g = 0.31$) than
those conducted in the classroom and the home ($g = 0.16$), the effect of 0.16 ($p = 0.18$)
represents the additive effect (i.e., value-add) of the informal learning environment. Although not
statistically significant, this is practically significant. It may be important for researchers,
practitioners, and parents to seriously consider the impact that all school-based interventions can
have with a home intervention component. School-based math interventions are effective
(Doabler et al., 2016; Powell & Fuchs, 2010), but perhaps home learning components that
support caregivers with training represent a critical step in closing achievement gaps.

**Implications of the Quality of Study Reporting**

As part of our ad hoc inclusion criteria, we excluded studies with methodology
limitations that may have impacted the results of the current meta-analysis. The methodological
and design decisions made by researchers can have a direct impact on the validity and accuracy
of estimated effects. For example, pre/post designs with no control group are susceptible to confounding maturation effects; Lipsey and Wilson (1993) show that one-group pre/post effect sizes are on average nearly twice as large as comparisons with control groups. Biased effect estimates due to methodological decisions will ultimately bias average effects in meta-analysis, reifying the idea of “garbage in, garbage out” (Borenstein et al., 2009). Methodological quality standards were enforced during the literature search process, so our quality review was specific to reporting quality for included studies (i.e., information that researchers provided in published reports of the intervention). High levels of study reporting quality are essential to enhance the generalizability, transferability, and replicability of intervention studies.

The results of this quality review (average QI rating of 0.59) were less favorable than the results of previous quality reviews of math interventions conducted in formal classroom settings (e.g., Author et al., Jitendra et al., 2021). Experts have suggested that study reporting quality may improve with the adoption of quality indicators across various fields (APA, 2008). Therefore, the current quality review offers valuable insights and recommendations for future research considering that researchers had not yet conducted an evaluation of study reporting quality with interventions conducted in informal learning environments. The current review of study reporting quality found that studies have improved in overall reporting since 2019 compared to earlier publications. Given that the evidence base on math interventions in informal learning environments is emerging, the increase in study reporting quality is a promising sign.

**Quality of Reporting Participant Information**

An aspect of reporting participant information is comparability of participants across conditions. Researchers frequently reported the use of random assignment and overall attrition. Reporting whether random assignment was used and the amount of attrition are key aspects for
evaluating the internal validity of a finding. Random assignment is important because the participants are more likely to be equal at baseline on confounders and therefore provide an unconfounded estimate of effect and having no attrition eliminates the possibility that the observed effect was due to the final sample containing only those participants that benefited most from the intervention (WWC, 2017). Studies less often met the desirable QI of reporting attrition across conditions beyond just reporting overall attrition.

Across studies, a relative strength of quality was authors’ reporting of the interventionist and their highest level of education. Providing information about interventionists allows the reader to determine whether intervention effects may have been impacted by differences between interventionists (Gersten et al., 2005). Although the results of the current study indicated that all interventionists were the child’s parent, future researchers may choose to examine the role of other types of caregivers. Therefore, we recommend that future researchers continue to provide detailed information about interventionists. Future research syntheses may examine other demographic variables that we did not consider, such as gender and linguistic diversity.

Although studies frequently reported child demographics including age, gender, race or ethnicity, and SES, studies less often reported whether children had a diagnosed disability, or if they were identified as dual-language learners. Previous research indicates that children from low SES backgrounds may be at a greater risk of experiencing math achievement gaps (Galindo & Sonnenschein, 2015), thus, the fact that so many authors presented SES information was expected. However, achievement trends also point to the gaps experienced by children who are dual language learners (NCES, 2019). A previous synthesis on school-based early math interventions reported smaller intervention effect sizes for treatment groups that had a larger proportion of participants identified as dual-language learners (Nelson & McMaster, 2019a). It is...
important for readers to understand how the results of included interventions transfer to other populations, including children with diverse learning needs. Moreover, it is critical that meta-analysts have access to full participant demographics in order to investigate variation in effects.

A limitation of the research base on early math interventions conducted in informal learning environments is the lack of representation of children with disabilities. When studies reported disability information, it was to inform readers that children with disabilities were excluded from the intervention (e.g., de Chambrier et al., 2019; Starkey & Klein, 2000). This was surprising, since research consistently reports on the early achievement gaps that exist for children with or at-risk of disabilities (Burchinal et al., 2011; Chu et al., 2019). Given the positive results of the current meta-analysis, early math interventions conducted in informal learning environments may be a fruitful opportunity to narrow achievement gaps for children who struggle with math. An extensive literature base reports on the effectiveness of school-based preschool to 3rd grade math interventions for students with disabilities or math difficulty (Doabler et al., 2016; Dyson et al., 2015; Fuchs et al., 2009; Powell & Fuchs, 2010; Siegler & Ramani, 2009). Yet, achievement gaps in math for many students, especially those who begin school below the 10th percentile (Morgan et al., 2009), remain or widen throughout elementary school (Nelson & Powell, 2018). Training caregivers to implement non-intensive interventions can provide children who struggle with the additional learning and practice opportunities they need to be proficient in math. Thus, we urge researchers to include or specifically target children with or at-risk of disabilities in their informal learning environment research and intervention studies. We also encourage researchers to report findings separately for these children to build a greater understanding of what works, for whom, and under what conditions.

Quality of Reporting Intervention and Comparison Conditions

http://mc.manuscriptcentral.com/rer
Studies need to present detailed information about the implementation of the intervention and the nature of the comparison condition (Gersten et al., 2005) to allow study replication with different populations. Replication is an essential component of the research cycle, especially when research in this field is still emerging (Makel & Plucker, 2014). Presenting this information also allows practitioners to bridge research to practice. Although we do not expect caregivers to access research studies to identify math activities to implement with children in the home, we expect trainers and organizations to access the information to determine best practices for supporting caregivers. Studies presented high-quality information related to interventionist training, recommended intervention dosage, and included math content and activities. These positive study quality results are similar to what previous researchers have reported for school-based math interventions (Jitendra et al., 2016).

An area of growth for this literature base is for authors to increase reporting related to the comparison condition and fidelity of implementation. Less than half of the time, studies reported the math content and activities used in the comparison condition, as well as the instructional time and nature of instruction in the comparison. Previous researchers have noted similar limitations of reporting fidelity of implementation (Author et al.; Jitendra et al., 2016; Liu et al., 2019). Providing researchers with information about the comparison condition is important in intervention research (including in meta-analyses of interventions) because the effect of the intervention condition is determined by comparing results to the comparison condition. The magnitude of this effect may be influenced by what instruction is provided in the comparison condition; for example, it may be important to know and control for what math activities caregivers in a business-as-usual control typically provide to children, as we can assume that most parents are providing some informal learning opportunities to their children (Susperreguy
& Davis-Kean, 2016). Future researchers may consider collecting and reporting more detailed information about the activities that occur in the comparison condition. Implementation fidelity is important from a research and practical standpoint. If interventions are found to be ineffective or to only produce negligible effects, researchers need to determine if the intervention was in fact, not effective, and requires modification, or if the intervention was simply not implemented as prescribed. Low fidelity of implementation can negatively influence expected treatment results (Hill & Erickson, 2019). Future researchers may review how studies measured fidelity of implementation in informal learning environments and determine if there are best practices for accurately capturing this information. Understanding more about how caregivers adhere to recommendations for interventions may reveal important barriers to fidelity and offer suggestions for improving the feasibility of interventions.

Quality of Reporting Outcome Measures

The effects reported from a meta-analysis on interventions rely on information derived from the outcome measures; thus, authors’ report of this information is desirable to conduct a high-quality review of the literature. Overall, studies met few (31%) QIs in this category; all but one of the QIs in this category were desirable indicators. The essential indicator representing whether authors administered the posttest measures immediately, within 2 weeks of the end of the intervention, was met only 30% of the time. This QI was met less often compared to school-based interventions (Jitendra et al., 2016; Lafay et al., 2019). However, it should be noted that all but two studies that did not meet this criterion simply did not report posttest administration time information at all. We encourage researchers to clearly report the time between the end of the intervention and the posttest administration so that readers are confident that the effects of the intervention are not due to competing hypotheses (Gersten et al., 2005).
INFORMAL MATH ENVIRONMENTS

We also promote delayed posttest data collection (or maintenance tests) so researchers can report on the long-term effects of interventions. Previous research on school-based interventions indicates that the effects of interventions fade beyond the prescribed intervention period (Bailey et al., 2020; Clements et al., 2013). It will be beneficial for researchers and practitioners to understand if fadeout effects also exist with interventions conducted in informal environments, and subsequently, researchers may consider investigating what level of ongoing support caregivers may benefit from to maintain intervention effects.

Gersten et al. (2005) also recommended the essential quality indicator of including both proximal and distal outcome measures; however, we focused this meta-analysis on math achievement measures that were aligned (proximal) to the intervention. Therefore, we removed this as an essential QI and deemed it as desirable. We recorded when studies administered non-achievement measures that captured the effect of the intervention on other child and caregiver aspects that may also be important to consider, such as the HME (e.g., caregiver reported scales, frequency counts of activities), child or caregiver attitude or motivation related to math, and frequency of “math talk” as observed by a researcher. It was a positive finding that more than half of studies included measures to capture the effect of the intervention on other aspects, especially given the evidence-base that demonstrates the relationship between the home math environment and children’s math achievement (Daucourt et al., 2021) and caregivers’ beliefs about the importance of math (Sonnenschein et al., 2012). We recommend future researchers continue to include non-achievement based measures in order to capture the all-encompassing effects of the intervention.

Other desirable indicators that researchers may report include reliability and validity information of the outcome measures. Our quality review results indicate reporting reliability...
and validity of measures is an area of improvement, similar to previous reviews of school-based
math interventions (Author et al.). It is important for readers to have this information because
tests that are used to determine the effectiveness of an intervention that also have low reliability
or validity represent a threat to the internal validity of the study. Moreover, when researchers use
unreliable assessments, the validity of the interpretations of the results of the intervention is
questionable (Weathington et al., 2010).

**Quality of Reporting Data Analysis and Results**

The quality of reporting related to the data analysis plan and results was a strength of the
included studies. Studies met 83% of the QIs in this category, and we encourage authors of
intervention studies to continue reporting information to support interpretation of the data
analysis plan and determination of intervention effectiveness. Previous reviews of math
interventions in school contexts reported inconsistent results for this category of QIs with some
studies reporting high scores (e.g., Jitendra et al., 2015) and others reporting low scores (e.g.,
Jitendra et al., 2016). One area of improvement for future research is for intervention studies to
report effect sizes, as researchers and practitioners may consider the reported effect size in
selecting the interventions that are likely to yield the greatest gains for children and families.

**Limitations**

As with any study, there are limitations of this meta-analysis. First, our results are limited
because only 23 studies (with 81 effect sizes) met inclusion. Although this reflects a developing
area of study, our power to detect meaningful effects, particularly in meta-regressions with many
predictors and interactions, was likely to be low. We also had strict inclusion criteria in order to
conduct a meta-analysis; there are math intervention studies conducted in informal learning
environments that were not represented in this meta-analysis (e.g., no math achievement
measure). Second, the results of this meta-analysis are limited in the generalizability of results to different populations of students. Readers should take caution in generalizing results to older children as the majority of studies in this meta-analysis included preschool-aged children.

Moreover, the results of this meta-analysis may not generalize to children with disabilities given that none of the included studies reported including children with or at-risk of disabilities. Few studies also reported inclusion of children identified as dual-language learners. Third, we were strict in our inclusion criteria regarding outcome measures. We considered only math achievement measures that were aligned to the purpose of the intervention. Future researchers may consider prioritizing the inclusion of aligned measures. Our meta-analysis also did not consider other outcome measures such as “math talk” or literacy outcomes (e.g., reading comprehension). Future meta-analyses may investigate other outcomes (e.g., “math talk”), including outcomes related to the environment (e.g., parent ratings of the HME). Finally, the majority of the studies included in this meta-analysis were conducted solely or partially in the home learning environment. Researchers have published math intervention studies conducted in other informal learning environments such as museums (e.g., Zippert et al., 2019) and grocery stores (e.g., Hanner et al., 2019); however, these studies did not meet inclusion criteria for a variety of reasons (e.g., design, outcome measures). The results of this meta-analysis may only generalize to interventions conducted, at least in part, in the home.

Conclusion

The results of this meta-analysis and quality review of early math interventions conducted in informal learning environments (viz., the home) broadly impact various education stakeholders by contributing to the evolution of this field, more specifically, fundamental understandings of how caregivers can support young children’s math learning. The results of this
meta-analysis report large effects (Kraft, 2020) on children’s math achievement outcomes. We offer several insights for future researchers and practitioners as they continue to identify best-practices for supporting children’s math learning with minimal training and cost. Regardless of the specific math content (e.g., counting, shape recognition) or activities (e.g., iPad apps, children’s books), the essential component of the intervention that researchers and practitioners should concentrate on is the degree of caregiver training and follow-up support. As previously explained, this training and follow-up support does not require extensive time or resources, offering broad access to the potential benefits for children of ensuring their caregivers are equipped to support them. The results of this study also offer encouraging findings related to the improvement of the quality of research reporting in recent years, which is promising for the generalizability, transferability, and replicability of research in this area. Important next steps in this line of research include expanding the knowledge base to include diverse populations of students and enhancing the level of reporting in studies as recommended by study quality guidelines.
References


INFORMAL MATH ENVIRONMENTS

in cardinal knowledge and executive function contribute to longer-term mathematical
https://doi.org/10.1016/j.jecp.2019.104668

knowledge and later school success. Teachers College Record, 115(6), 1-29.
https://psycnet.apa.org/record/2013-18273-005

scale-up model for teaching mathematics with trajectories and technologies: Persistence
https://doi.org/10.3102%2F0002831212469270

Erlbaum.

Cook, B. G., & Cook, L. (2017). Do research findings apply to my students? Examining study
samples and sampling. Learning Disabilities Research & Practice, 32(2), 78-84.
https://doi.org/10.1111/ldr.12132

and meta-analysis (2nd ed.). Russell Sage Foundation.

environment and math achievement: A meta-analysis. Psychological Bulletin, 147(6),
565–596. https://doi.org/10.1037/bul0000330

start: Early numeracy mastery predicts high school math course-taking and college
INFORMAL MATH ENVIRONMENTS


https://doi.org/10.1521/scpq.2005.20.4.371


https://doi.org/10.1007/s11162-013-9322-7


https://doi.org/10.1177/0956797617729817


Routledge.


https://doi.org/10.1080/17482798.2019.1613246


[http://dx.doi.org/10.1136/bmjopen-2015-010247](http://dx.doi.org/10.1136/bmjopen-2015-010247)

[https://doi.org/10.1177/0014402920969187](https://doi.org/10.1177/0014402920969187)

[https://doi.org/10.1177/0022219413487408](https://doi.org/10.1177/0022219413487408)

[https://doi.org/10.1177/0014402915625062](https://doi.org/10.1177/0014402915625062)


[https://doi.org/10.1177%2F0731948718823083](https://doi.org/10.1177%2F0731948718823083)


Lee, J. (2012). College for all: Gaps between desirable and actual P–12 math achievement
INFORMAL MATH ENVIRONMENTS


https://doi.org/10.3102/0013189X11432746


http://dx.doi.org/10.1037/a0014532


INFORMAL MATH ENVIRONMENTS


Passolunghi, M. C., & Lanfranchi, S. (2012). Domain-specific and domain-general precursors of


Pustejovsky, J. (2019). clubSandwich: Cluster-Robust (sandwich) variance estimators with
small-sample corrections. https://CRAN.R-project.org/package=clubSandwich


INFORMAL MATH ENVIRONMENTS


Table 1

*Summary of Coding Protocol*

<table>
<thead>
<tr>
<th>Section of Coding Protocol</th>
<th>Explanation of Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Study Information and Participant Demographics</td>
<td>Authors, year, journal, type of publication, and location</td>
</tr>
<tr>
<td>Methodological Information</td>
<td>Study design, number of treatment and control groups, assignment procedures, treatment fidelity, quality of implementation, and interrater agreement information</td>
</tr>
<tr>
<td>Intervention Features</td>
<td>Interventionist, intervention location, interventionist training and follow-up support characteristics and time, recommended and reported intervention implementation time (e.g., number of sessions per week, total weeks, number of minutes of play), intervention activities (e.g., games, books, puzzles), and mathematics intervention content (e.g., number identification, counting, basic facts)</td>
</tr>
<tr>
<td>Outcome Measures</td>
<td>Outcome measures administered to measure the effectiveness of the intervention</td>
</tr>
<tr>
<td>Effect Sizes</td>
<td>Means and standard deviations of pre- and posttest mathematics achievement measures, sample size, and design</td>
</tr>
</tbody>
</table>

Information about the control condition including if the control was described and if so what types of activities and content were implemented.

Outcome measures other than mathematics achievement (e.g., working memory, literacy, HME scales) Name, citation, respondent, content focus, administration time, post-test administration window, number of items, reliability, validity, and publication type (e.g., norm-referenced, researcher developed)

Information to calculate variance including the pre-test/post-test correlation when provided

Quality of the data analysis plan, whether authors reported effect sizes, and whether results were reported clearly
### Section of Coding Protocol

<table>
<thead>
<tr>
<th>Study Reporting Quality</th>
<th>Explanation of Codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 Quality of participant description: Captured whether studies provided sufficient detail related to child and interventionist demographics and use of random assignment</td>
<td></td>
</tr>
<tr>
<td>2.0 Quality of the description of intervention and comparison conditions: Identified if studies clearly described the intervention and control conditions with relevant detail to support readers’ understanding of the intervention procedures, including activities, mathematics content, and dosage; Captured study procedures for measuring and reporting implementation fidelity</td>
<td></td>
</tr>
<tr>
<td>3.0 Quality of outcome measures: Coded whether studies reported administering outcome measures at immediate post-test; Coded if measures were administered beyond immediate post-test to determine the long-term effects of the intervention; Determined if studies reported administering outcome measures that captured the effect of the intervention on non-academic skills that may be especially relevant for informal learning environments (e.g., math talk, home math environment scales, parent involvement)</td>
<td></td>
</tr>
<tr>
<td>4.0 Quality of the description of data analysis and results: Coded whether authors described an appropriate data analysis plan; Coded whether authors reported important links between the data analysis plan and research questions; Recorded if the results were clearly articulated; Recorded if the authors reported an effect size</td>
<td></td>
</tr>
</tbody>
</table>

Note. A complete draft of the coding protocol that was used for this study can be found at (INSERT LINK UPON ACCEPTANCE TO OPEN ACCESS CODE BOOK).
a. Only information from math achievement outcomes measures was captured and used to calculate effect sizes.
Table 2

Tests of Residual Variability and Omnibus Tests of Moderators

<table>
<thead>
<tr>
<th>Variable</th>
<th>k</th>
<th>N</th>
<th>Effect Size</th>
<th>Q(df)</th>
<th>p</th>
<th>Omnibus Test</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome measure content</td>
<td>23</td>
<td>81</td>
<td></td>
<td>Qf(72) = 95.69</td>
<td>0.03</td>
<td>F(9,72) = 1.71</td>
<td>0.10</td>
</tr>
<tr>
<td>Intensity of caregiver training</td>
<td>23</td>
<td>81</td>
<td></td>
<td>Qf(78) = 94.67</td>
<td>0.10</td>
<td>F(2,78) = 3.04</td>
<td>0.05</td>
</tr>
<tr>
<td>Intervention location</td>
<td>23</td>
<td>81</td>
<td></td>
<td>Qf(79) = 102.95</td>
<td>0.03</td>
<td>F(1,79) = 1.62</td>
<td>0.21</td>
</tr>
<tr>
<td>Number of sessions</td>
<td>18</td>
<td>70</td>
<td></td>
<td>Qf(68) = 94.03</td>
<td>0.02</td>
<td>F(1,68) = 0.20</td>
<td>0.66</td>
</tr>
<tr>
<td>Length of intervention</td>
<td>21</td>
<td>79</td>
<td></td>
<td>Qf(77) = 103.97</td>
<td>0.02</td>
<td>F(1,77) = 0.04</td>
<td>0.85</td>
</tr>
<tr>
<td>Main intervention activity</td>
<td>23</td>
<td>81</td>
<td></td>
<td>Qf(75) = 98.99</td>
<td>0.03</td>
<td>F(5,75) = 0.49</td>
<td>0.78</td>
</tr>
<tr>
<td>Intervention math content</td>
<td>23</td>
<td>81</td>
<td></td>
<td>Qf(74) = 95.84</td>
<td>0.04</td>
<td>F(6,74) = 1.33</td>
<td>0.26</td>
</tr>
</tbody>
</table>
### Table 3
Summary of Individual Moderator Results

<table>
<thead>
<tr>
<th>Moderator</th>
<th>$k$</th>
<th>$N$</th>
<th>Effect Size</th>
<th>$g$</th>
<th>$SE$</th>
<th>$df$</th>
<th>$p$</th>
<th>95% CI LL</th>
<th>95% CI UL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome measure content</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Counting</td>
<td>10</td>
<td>22</td>
<td>0.20</td>
<td>0.15</td>
<td>11.14</td>
<td>0.22</td>
<td>-0.14</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>Numeral Identification</td>
<td>6</td>
<td>12</td>
<td>0.17</td>
<td>0.21</td>
<td>8.88</td>
<td>0.44</td>
<td>-0.30</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>Addition</td>
<td>4</td>
<td>6</td>
<td>0.18</td>
<td>0.21</td>
<td>7.10</td>
<td>0.41</td>
<td>-0.30</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>Quantity discrimination</td>
<td>5</td>
<td>8</td>
<td>0.15</td>
<td>0.24</td>
<td>0.62</td>
<td>0.21</td>
<td>-0.38</td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Number line estimation</td>
<td>3</td>
<td>6</td>
<td>-0.08</td>
<td>0.24</td>
<td>5.75</td>
<td>0.74</td>
<td>-0.67</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Missing number or number order</td>
<td>3</td>
<td>3</td>
<td>0.07</td>
<td>0.29</td>
<td>5.27</td>
<td>0.82</td>
<td>-0.67</td>
<td>0.81</td>
<td></td>
</tr>
<tr>
<td>Non-numerical skills$^a$</td>
<td>6</td>
<td>10</td>
<td>0.38</td>
<td>0.21</td>
<td>8.91</td>
<td>0.10</td>
<td>-0.09</td>
<td>0.85</td>
<td></td>
</tr>
<tr>
<td>Comprehensive early math$^b$</td>
<td>6</td>
<td>7</td>
<td>0.30</td>
<td>0.13</td>
<td>4.02</td>
<td>0.07</td>
<td>-0.05</td>
<td>0.65</td>
<td></td>
</tr>
<tr>
<td>Comprehensive early numeracy$^c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Intensity of caregiver training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No training</td>
<td>4</td>
<td>12</td>
<td>-0.02</td>
<td>0.06</td>
<td>3.11</td>
<td>0.76</td>
<td>-0.21</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>Initial training session only</td>
<td>5</td>
<td>25</td>
<td>0.07</td>
<td>0.07</td>
<td>3.73</td>
<td>0.34</td>
<td>-0.11</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>Initial training and follow-up support</td>
<td>15</td>
<td>44</td>
<td>0.43</td>
<td>0.12</td>
<td>10.29</td>
<td>0.01</td>
<td>0.15</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td><strong>Intervention location</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child’s home only</td>
<td>19</td>
<td>68</td>
<td>0.32</td>
<td>0.10</td>
<td>16.20</td>
<td>0.01</td>
<td>0.11</td>
<td>0.54</td>
<td></td>
</tr>
<tr>
<td>School intervention with home aspect</td>
<td>5</td>
<td>13</td>
<td>0.16</td>
<td>0.09</td>
<td>4.11</td>
<td>0.18</td>
<td>-0.11</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td><strong>Number of sessions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td>0.29</td>
<td>0.10</td>
<td>12.89</td>
<td>0.01</td>
<td>0.08</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>Sessions</td>
<td>18</td>
<td>70</td>
<td>0.001</td>
<td>0.003</td>
<td>2.16</td>
<td>0.80</td>
<td>-0.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td><strong>Length of intervention</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td>0.27</td>
<td>0.10</td>
<td>15.10</td>
<td>0.02</td>
<td>0.06</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Weeks of Intervention</td>
<td>21</td>
<td>79</td>
<td>0.002</td>
<td>0.01</td>
<td>2.30</td>
<td>0.82</td>
<td>-0.03</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td><strong>Main intervention activity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Children’s math story books</td>
<td>3</td>
<td>4</td>
<td>0.22</td>
<td>0.05</td>
<td>1.97</td>
<td>0.04</td>
<td>0.01</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>Number board games</td>
<td>3</td>
<td>26</td>
<td>0.24</td>
<td>0.11</td>
<td>2.95</td>
<td>0.12</td>
<td>-0.11</td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>Card games or dice games$^d$</td>
<td>5</td>
<td>21</td>
<td>0.53</td>
<td>0.32</td>
<td>3.68</td>
<td>0.18</td>
<td>-0.39</td>
<td>1.45</td>
<td></td>
</tr>
<tr>
<td>Varying activities with options$^e$</td>
<td>6</td>
<td>14</td>
<td>0.14</td>
<td>0.11</td>
<td>4.10</td>
<td>0.26</td>
<td>-0.15</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Food Routines</td>
<td>2</td>
<td>2</td>
<td>0.23</td>
<td>0.22</td>
<td>1.00</td>
<td>0.48</td>
<td>-2.52</td>
<td>2.99</td>
<td></td>
</tr>
<tr>
<td>Formal caregiver curriculum$^f$</td>
<td>5</td>
<td>14</td>
<td>0.23</td>
<td>0.17</td>
<td>3.46</td>
<td>0.27</td>
<td>-0.29</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td><strong>Intervention math content</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number Skills</td>
<td>19</td>
<td>72</td>
<td>0.22</td>
<td>0.12</td>
<td>16.51</td>
<td>0.07</td>
<td>-0.02</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>Relations Skills</td>
<td>16</td>
<td>47</td>
<td>0.26</td>
<td>0.10</td>
<td>16.40</td>
<td>0.03</td>
<td>0.04</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>Operations Skills</td>
<td>12</td>
<td>36</td>
<td>0.32</td>
<td>0.13</td>
<td>8.33</td>
<td>0.04</td>
<td>0.03</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>Measurement/Geometry/Spatial</td>
<td>9</td>
<td>18</td>
<td>0.50</td>
<td>0.21</td>
<td>5.65</td>
<td>0.05</td>
<td>-0.01</td>
<td>1.01</td>
<td></td>
</tr>
<tr>
<td>Organizing/Sorting</td>
<td>5</td>
<td>9</td>
<td>0.16</td>
<td>0.16</td>
<td>3.5</td>
<td>0.39</td>
<td>-0.31</td>
<td>0.63</td>
<td></td>
</tr>
<tr>
<td>Sequencing/Patterns</td>
<td>7</td>
<td>15</td>
<td>0.30</td>
<td>0.12</td>
<td>5.24</td>
<td>0.06</td>
<td>-0.01</td>
<td>0.61</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Given that some studies included more than one treatment condition, not all categories sum to 23 studies.

a. Non-numerical skills include skills such as measurement, shape recognition, and patterning.
b. Comprehensive early math skills refer to multiple math skills above and beyond numeracy skills (e.g., TEMA).
c. Comprehensive early numeracy skills refer to measures that include multiple early numeracy skills (e.g., counting, comparison, numeral identification, simple addition).
d. The main activity was card or dice games only and were specifically recommended for caregivers (e.g., number card or dice games with explicit instructions).
e. The main activity comprised several varying counting or number games or activities with some caregiver choice in implementation over the course of the intervention.
f. The main activity was a specified set of activities and sequence of lessons for caregivers to implement.

http://mc.manuscriptcentral.com/rer
INFORMAL MATH ENVIRONMENTS

Table 4

Summary of Percentage of Quality Indicators Met Across Categories, Subcategories, and Individual Codes

<table>
<thead>
<tr>
<th>Quality Indicator category, subcategories, and individual codes</th>
<th>M  (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 Participant description</td>
<td></td>
</tr>
<tr>
<td>1.1 Child demographics</td>
<td></td>
</tr>
<tr>
<td>1.1.1 Did authors report child age or grade?</td>
<td>.66 (.16)</td>
</tr>
<tr>
<td>1.1.2 Did authors report child gender?</td>
<td>.61 (.17)</td>
</tr>
<tr>
<td>1.1.3 Did authors report child dual language learner status?</td>
<td>1.0 (0.0)</td>
</tr>
<tr>
<td>1.1.4 Did authors report child race or ethnicity?</td>
<td>.87 (.34)</td>
</tr>
<tr>
<td>1.1.5 Did authors report child disability information?</td>
<td>.09 (.29)</td>
</tr>
<tr>
<td>1.1.6 Did authors report family socioeconomic status?</td>
<td>.78 (.42)</td>
</tr>
<tr>
<td>1.1.7 Did authors report languages spoken in the family’s home?</td>
<td>.26 (.45)</td>
</tr>
<tr>
<td>1.2 Interventionist demographics</td>
<td></td>
</tr>
<tr>
<td>1.2.1 Did authors report the interventionist (e.g., parent, sibling)?</td>
<td>.83 (.24)</td>
</tr>
<tr>
<td>1.2.3 Did the authors report the interventionists highest level of education?</td>
<td>1.0 (0.0)</td>
</tr>
<tr>
<td>1.3 Comparability of participants across conditions</td>
<td></td>
</tr>
<tr>
<td>1.3.1 Did the authors report using random assignment?</td>
<td>.61 (.28)</td>
</tr>
<tr>
<td>1.3.2 Did the authors report overall attrition from pre to post-test?</td>
<td>.78 (.49)</td>
</tr>
<tr>
<td>1.3.3 Did the authors report differential attrition across conditions?</td>
<td>.74 (.42)</td>
</tr>
<tr>
<td>2.0 Description of intervention and comparison conditions</td>
<td></td>
</tr>
<tr>
<td>2.1 Intervention condition was clearly described</td>
<td>.73 (.41)</td>
</tr>
<tr>
<td>2.1.1 Did the authors report whether interventionists received training?</td>
<td>.81 (.39)</td>
</tr>
<tr>
<td>2.1.2 Did the authors report the intervention length (weeks, months)?</td>
<td>.91 (.29)</td>
</tr>
<tr>
<td>2.1.3 Did the authors report the intervention intensity (sessions/week)?</td>
<td>.78 (.42)</td>
</tr>
<tr>
<td>2.1.4 Did the authors report the intervention session length (minutes)?</td>
<td>.35 (.49)</td>
</tr>
<tr>
<td>2.1.5 Did the authors report the intervention activities with replicability?</td>
<td>.78 (.42)</td>
</tr>
<tr>
<td>2.1.6 Did the authors report the intervention content with replicability?</td>
<td>.72 (.45)</td>
</tr>
<tr>
<td>2.2 Comparison condition was clearly described</td>
<td>.42 (.46)</td>
</tr>
<tr>
<td>2.2.1 Did the authors report content or activities in the comparison condition?</td>
<td>.48 (.51)</td>
</tr>
<tr>
<td>2.2.2 Did the authors report instructional time (either length, sessions, or session length) in the comparison condition?</td>
<td>.48 (.51)</td>
</tr>
<tr>
<td>2.2.4 Did the authors provide any documentation of the nature of instruction in the comparison condition?</td>
<td>.30 (.47)</td>
</tr>
<tr>
<td>2.3 Fidelity of implementation</td>
<td>.49 (.35)</td>
</tr>
<tr>
<td>Quality Indicator category, subcategories, and individual codes</td>
<td>M</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------------------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>2.3.1 Did the authors report that they measured fidelity of implementation?</td>
<td>.71</td>
</tr>
<tr>
<td>2.3.2 Did the authors report the results of the fidelity of implementation?</td>
<td>.70</td>
</tr>
<tr>
<td>2.3.3 Did the authors report that they measured fidelity of the quality of implementation (i.e., reporting fidelity beyond adherence to procedures and recommended dosage)?</td>
<td>.17</td>
</tr>
</tbody>
</table>

3.0 Outcome Measures                                                                                                      | .31   | (.18) |
| 3.1 Multiple measures and administration                                                                                   | .32   | (.26) |
| 3.1.1 Did the authors report administering non-achievement measures (e.g., home numeracy scales, math motivation or interest questionnaires, parental involvement, math talk) in addition to achievement measures to generalize the effectiveness of the intervention to the child’s home environment? | .56   | (.51) |
| 3.1.2 Did the authors report that they administered the post-test immediately, within 2 weeks of the end of the intervention? | .30   | (.47) |
| 3.1.3 Did the authors report that they also administered delayed post-tests to determine the long-term effects of the intervention? | .09   | (.29) |

3.2 Reliability and validity                                                                                                | .30   | (.25) |
| 3.2.1 Did the authors report the internal reliability of the outcome measures?                                               | .43   | (.51) |
| 3.2.2 Did the authors report other, additional, reliability estimates (e.g., test-rest, alternate form) for the outcome measures? | .35   | (.49) |
| 3.2.3 Did the authors report information to support the validity of the outcome measures (e.g., concurrent or predictive validity, expert validity)? | .30   | (.47) |
| 3.2.4 Did the authors report that the data collectors were blind to study conditions?                                         | .22   | (.42) |
| 3.2.5 Did the authors report inter-scorer agreement for the outcome measures?                                                   | .22   | (.42) |

4.0 Data analysis and results                                                                                                                                                   | .83   | (.15) |
| 4.1 Data analysis plan described and appropriate                                                                           | .91   | (.18) |
| 4.1.1 Did the authors describe the data analysis plan?                                                                      | .96   | (.21) |
| 4.1.2 Did the authors appropriately link the data analysis plan to research questions?                                     | .83   | (.39) |
| 4.1.3 Did the authors appropriately link the data analysis to the unit of analysis?                                         | .96   | (.21) |
| 4.2 Results                                                                                                               | .72   | (.25) |
| 4.2.1 Did the authors clearly and coherently report results of the intervention to convey the effectiveness of the intervention? | 1.0   | (.00) |
| 4.2.2 Did the authors report effect sizes for the math outcomes?                                                             | .43   | (.51) |

Note. Individual codes that are in italics represent desirable indicators; all other indicators are essential indicators.
Figure 1

PRISMA Diagram Illustrating the Literature Search Procedures

Notes. The figure is a PRISMA Diagram that describes the literature search procedures that were used to exclude abstracts and studies from inclusion in this meta-analysis.
Figure 2

Funnel and Contour Plots

Note. This figure shows the results of the funnel plot (left) and contour-enhanced funnel plot (right) that were used to examine publication bias. The funnel plot shows observed effect sizes to their standard errors. Effect sizes located in the gray region were further scrutinized and ultimately retained. We used the contour funnel plot to evaluate publication bias resulting from suppression of non-significant findings.