8-2022

Early Math Interventions in Informal Learning Environments: Technical Report

Gena Nelson
University of Oregon

Hannah Carter
Boise State University

Peter Boedeker
Baylor College of Medicine

Early Math Interventions in Informal Learning Environments: Technical Report

Gena Nelson
University of Oregon

Hannah Carter
Boise State University

Peter Boedeker
Baylor College of Medicine

The purpose of this technical report is to document the procedures of the National Science Foundation (NSF) funded grant (DUE #2000292), *Investigating Effective Methods that Adults Use to Improve Children's Math Achievement in Informal Learning Environments: A Meta-Analysis.* Specifically, this technical report documents the method and results of the literature search process, abstract screening, and full text-review. In addition, Appendix A includes our abstract screening protocol and Appendix B includes our full-text review protocol. This technical report also describes the article coding procedures and all the methods that senior personnel used to train graduate assistants (GAs) about each phase of the meta-analysis grant. This document represents the work of the first year of a 3-year meta-analysis project.

The main purpose of this project is to conduct a meta-analysis of math interventions conducted in informal learning environments with math achievement outcomes. However, our literature search process revealed several additional research questions that are worth exploring (e.g., studies that have "math talk" outcomes, studies that have "home numeracy environment" outcomes). Therefore, it should be noted that our literature review process was broader than only identifying studies with math achievement outcomes as a method to prepare for future manuscript submissions.

Method

Inclusion and Exclusion Criteria

We used the following inclusion and exclusion criteria throughout the literature search process for inclusion in the meta-analysis:

1. The study investigated the effects of a math intervention implemented in an informal learning environment. Informal learning environments were defined as home, out of school programs held at community spaces such as libraries and museums, and everyday experiences (Gerber et al., 2001). Studies conducted only in school settings were excluded. Studies that were conducted in school environments that had home components were included if the effects of the home component could be isolated from the school-based intervention; if we were not able to isolate the effects of the home component these studies were included in the quality review study only and were excluded from the meta-analysis.
2. Children in the study were administered at least one math achievement outcome measure to determine effectiveness. Studies that only administered measures of "math talk" were excluded from the main meta-analysis focused on math achievement outcomes. These studies however, were retained for the quality review study and a follow-up meta-analysis focused on "math talk" outcomes will be conducted in the future. Studies that only administered cognitive measures that are not math *achievement* measures (e.g., executive functioning, SFON) were excluded from the main meta-analysis focused on math achievement outcomes.
3. The primary implementer or facilitator of the math program was a caregiver, including parents or guardians, adults who play an informal childcare role (e.g., grandparent, neighbor, older sibling), and afterschool, library, or museum staff. Studies conducted with researchers or teachers as the main facilitator of the program were excluded.
4. Participants were children 3 years, 0 months (average age at the start of the intervention) to the end of third grade (average age less than 9 years old). Studies with children

younger than an average of 3 years old were excluded unless data were disaggregated for children within the target age. Studies with children older than 9 years old were excluded unless data were disaggregated for children within the target age.

5. The study has an experimental or quasi-experimental group design. Studies that used single case design or qualitative methods were excluded. Studies that used post-test design only methods without randomization were excluded. Studies that included only a treatment group were also excluded.

5. Appropriate information to calculate effect sizes (e.g., means, *SD*s, F statistics, *t*-tests) was reported. Studies that did not report correct information to calculate effect sizes were included if authors were able to provide follow-up information; all authors of studies with missing information were contacted by the first author.

6. Results were reported in English.

Literature Search

We took several steps to conduct a comprehensive review of the literature. We did not restrict our literature search by publication year or publication type; peer-reviewed journal articles, dissertations, conference presentations, book chapters, and technical reports were all considered for inclusion in this meta-analysis. Our review of literature included: a.) electronic database searches; b.) ProQuest Dissertations & Theses database searches; c.) electronic table of contents review for relevant journals not captured elsewhere; d.) direct contacts to relevant experts for published studies; e.) messages in relevant organizational email listservs; f.) pre-print searches in PsyArXiv (https://psyarxiv.com/); g.) reference list searches; and h.) forward citation searches.

Electronic Database Search

First, we searched electronic databases including: Academic Search Premier, Education Research Complete, ERIC, PsycARTICLES, and PsycINFO. The following Boolean search string was used: (intervention OR activity OR training OR tutoring) AND (math* OR numeracy OR "number sense" OR “math* play” OR “math talk” OR “book reading” OR tablet OR “e-book” OR geometry OR algebra OR “informal learning”) 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”).

ProQuest Dissertations & Theses Database Search

Second, although we did not restrict any of the above databases by publication type (meaning that we also captured technical reports and dissertations in the search described above), we also searched the ProQuest Dissertations & Theses databases to identify any theses or dissertations that we may have missed with the initial search. We acquired more than 320,000 dissertations matching our search terms, which could be a result of 1. ProQuest Dissertations & Theses capturing dissertations from many non-education databases and 2. the documents are generally lengthier than a peer-reviewed journal (meaning there is more text to match with search terms). This number was substantially greater than the number of articles that matched our search from the other five electronic databases. Upon scanning the titles of approximately 100 of the dissertations we captured, many of the titles suggested that the studies were not early mathematics intervention studies. For that reason, we restricted our search to dissertations with Boolean search string terms appearing in the abstract only.

Table of Contents Review

Third, we searched the electronic table of contents for relevant journals that had not appeared in the electronic database search or occurred with low frequency and were journals that experts in the areas of early mathematics, informal learning, and meta-analysis recommended reviewing. We used the same Boolean search string 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*. In addition to using this method to search relevant journal table of contents, we also conducted a hand search of journals that did not allow for online electronic search using the Boolean search string and that experts in the field deemed highly relevant to the purpose of our systematic review, including five journals: *Children and Libraries*; *Cognitive Development*; *Early Childhood Research Quarterly*; *Journal of Experimental Child Psychology*; and *Journal of Research in Mathematics Education*. The hand search was completed by accessing the full table of contents for these journals for the publication years 2016 to 2021. All titles and abstracts were searched for all volumes and issues for these publication years.

Direct Communication with Experts in the Field

Fourth, experts (mainly our advisory board members and consultants) in the areas of early mathematics, informal learning, and meta-analysis provided us with a list of researchers to contact directly for research studies they conducted related to the topic of this meta-analysis. We contacted 20 researchers about the purpose of our study and provided a set of broad inclusion and exclusion criteria for our meta-analysis. We included criteria that were broader than our specific criteria for this meta-analysis so that we could appropriately review the full-text of any articles that we received. The criteria that we sent researchers was the following: (a) the study investigates the effects of a math intervention implemented in an informal learning environment; (b) the child participants were administered at least one math outcome measure to determine the effectiveness of the intervention; (c) the study includes at least some participants who are children between the ages of 3 years, 0 months (average start age) to the end of third grade (approximately 9 years old); (d) the study used a group design; (e) the study provides information to calculate effect sizes (e.g., means, SDs, F statistics, t-tests), or authors are able to provide this information upon request; and, (f) the study is available in English. The list of researchers we contacted included: Drs. Martha Alibali, Elizabeth Brannon, Julie Booth, Doug Clements, Bert DeSmedt, Sarah Eason, Lisa Feigenson, Susan Levine, Melissa Libertus, Gigliana Melzi, Kelly Mix, Frank Niklas, John Opfer, David Purpura, Geetha Ramani, Julie Sarama, Anna Shusterman, Susan Sonnenschein, Colleen Uscianowski, and David Uttal.

Contacting Relevant Organizations

Fifth, experts (mainly our advisory board members and consultants) in the areas of early mathematics, informal learning, and meta-analysis provided us with a list of organizations that maintain an email listserv with subscribers who may have conducted research studies that are aligned to the focus of this meta-analysis. Although we contacted several organizations, the following are organizations that allowed us to post an announcement to the listserv as non-members: 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 a contact at the National Science Foundation's Center for the Advancement of Informal Science Learning for information on any former grant projects that may have had research studies aligned to the purpose of this meta-analysis.

Pre-print Publications

Sixth, we searched PsyArXiv (https://psyarxiv.com/) for any pre-prints of studies that authors may have submitted to the repository. Based on the different electronic search function of this database, we searched general terms including: "math AND intervention" "home AND math" "home AND intervention" "home AND numeracy" "numeracy AND intervention" "tutoring AND math" "home AND tutoring" "informal AND math" "activity AND math" "activity AND numeracy" "play AND math" "play AND numeracy." Across these broad search terms, there were 94 total pre-prints and none of which we deemed eligible for our meta-analysis. Both the first and second authors independently conducted the search of pre-prints.

Reference List Search of Included Studies

Seventh, we conducted a search of reference lists of included articles. We did not search the reference lists of all included studies; instead, we selected three included studies that were currently under review at journals as they represented the most recent research on this topic. These studies were also deemed methodologically sound for inclusion in the meta-analysis. We ensured that the three studies we selected represented different types of interventions in informal learning environments, including a text messaging intervention (Napoli & Purpura, under review), a picture book intervention (Purpura et al., under review), and a food routine intervention (Leyva et al., under review). We searched the reference lists and created a catalogue of any studies that were not already captured in the other literature search processes. The GAs were advised to start at the beginning of the reference list and check every reference against the results of the initial electronic database search. Any citation in the reference list that was not also included in the main database results was added to a new Excel database for abstract review. The GAs then reviewed these abstracts in the same manner as all other abstracts (described below).

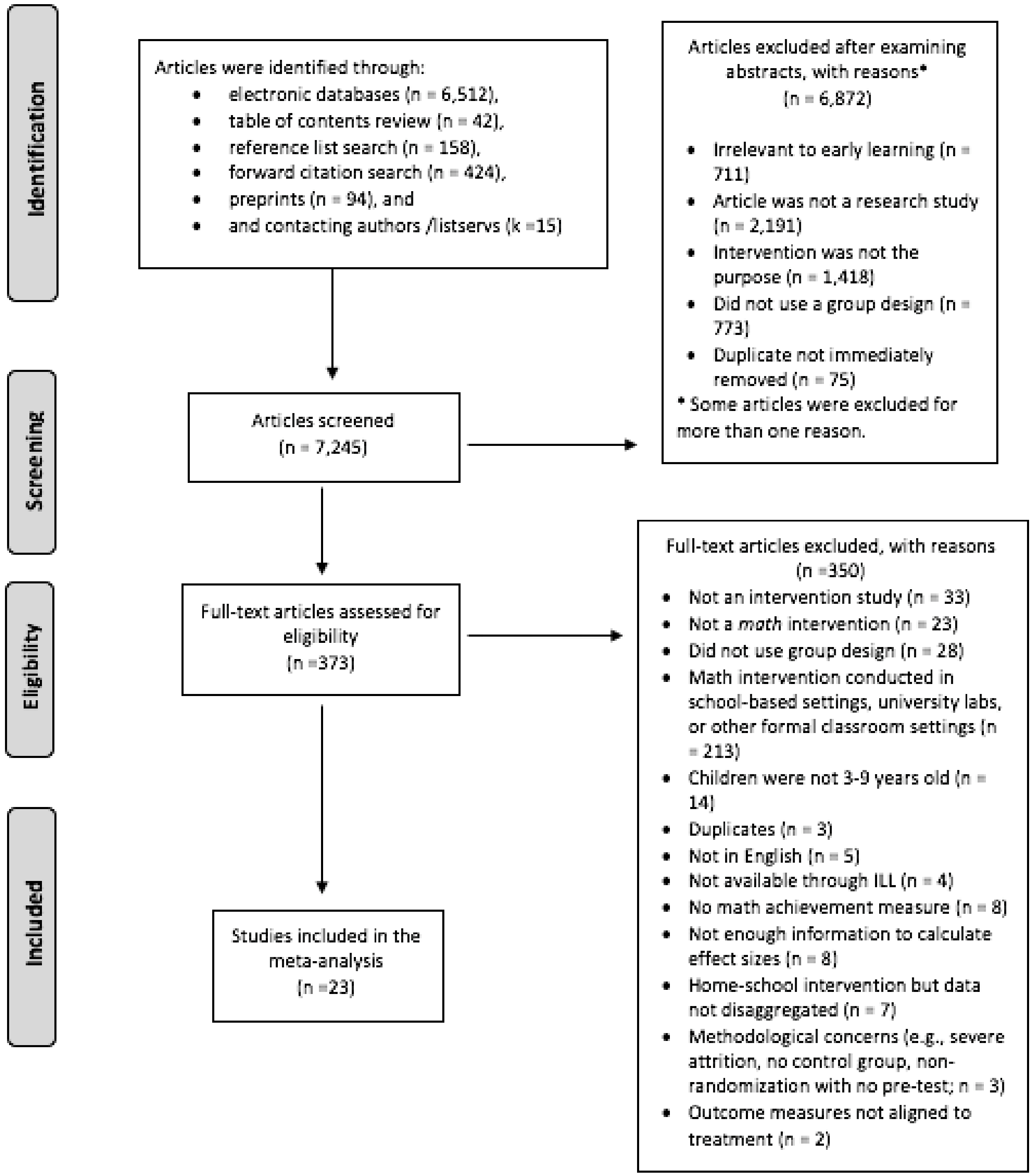
Forward Citation Search

Finally, we conducted a forward citation search of three publications (Niklas et al., 2016; Starkey & Klein, 2000; Vandermaas-Peeler et al., 2012). To identify the three publications we used to conduct the forward citation search, we first entered all of the included studies into Google Scholar and used the "cited by" function to identify how many times an article had been cited by other publications. We identified the three studies with the most "cited by" publications. Then, we examined the list of articles that appeared and checked every citation against the results of the initial electronic database search. Any citation in the forward citation list that was not also included in the main database results was added to a new Excel database for abstract review.

Literature Search Results

The first author led each literature search procedure; she has experience organizing and conducting literature searches, systematic reviews, and meta-analyses. We captured the following number of studies from each of the search methods: electronic database search including dissertations (k = 6,512); table of contents review (k = 42); contacting authors and organizational listservs (k = 15); preprint search (k = 94); reference list search of included articles (k = 158); and a forward citation search of a select subset of included studies (k = 424). In total, we reviewed 7,245 abstracts; 100% of abstracts were double screened. From the abstract screening phase, we identified 373 studies for full-text review; 100% of articles were double coded for full-text review. Upon completion of full-text review, we identified 23 studies for inclusion in this meta-analysis with a total of 81 effect sizes. The processes for abstract screening and full-text review are described in detail below; however, the results of abstract screening and full-text review are summarized in the PRISMA diagram below.

Figure 1. PRISMA Diagram



Abstract Screening and Training

The first and second author trained three GAs to complete the abstract screening. See Appendix A for the abstract screening protocol. The GAs were pursuing master's degrees in Counseling, Early and Special Education, and Elementary Education. One of the GAs had previous experience conducting literature reviews. To ensure accurate identification of studies to include this in this meta-analysis, training consisted of several components. The abstract review process began in January 2021; however, training began in October 2020.

During four 1.5-hour meetings, the first and second author trained the GAs via Zoom. In addition to team building exercises, the trainings focused on discussion and practice related to: 1. the purpose of the meta-analysis; 2. defining features of meta-analyses, 3. features of published early math intervention research studies; and 4. best practices for abstract screening. Before each of the four 1.5-hour meetings, GAs were assigned a reading (e.g., Cheung & McBride-Chang, 2015; Polanin, 2018; Therrien et al., 2020), as well as small tasks to prepare for discussion (e.g., completing a review of abstracts using set criteria from an unrelated project).

At the conclusion of the four 1.5-hour meetings, and once the GAs demonstrated an understanding of meta-analysis, purpose of the project, and the general abstract review process, the first and second author implemented a 2-hour training. This training included a review of the abstract review criteria specifically 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 10 abstracts. After the 2-hour training, the GAs were assigned the same 25 abstracts to review before the 1-hour follow-up meeting the next day. During the 1-hour follow-up training, the GAs, first, and second author reviewed abstracts where there was any disagreement about inclusion. The GAs also provided the first and second author with feedback about the abstract screening tool and database and overall process. The first author refined the abstract screening tool. See Table 1 for a table of Kappa values that represent the level of agreement using the abstract screening tool throughout the abstract screening process.

After discussing the first 25 practice articles and making refinements to the screening tool, the GAs each coded the same 150 articles. Agreement for the inclusion of articles increased from the first set of 25 articles. Again, as a group we discussed discrepancies and made refinements to the abstract screening tool (e.g., we added criteria to support GAs in excluding irrelevant studies, we added examples). For the next stage of training, we reviewed abstracts that had disagreement across any of the raters for the first 175 studies. The first author created an Excel database with a list of all studies, (k = 65). Then, each graduate student re-coded each abstract as if it was the first time they were coding the abstract. Agreement on these 65 re-coded articles was .49. The GAs then met separately, without the first or second author present, to review the remaining discrepancies and come to a consensus about the final code (include, exclude) for those abstracts.

The remaining abstracts were coded by only two GAs (as opposed to all abstracts being coded by three GAs). Each graduate student coded between 400 and 600 abstracts each week. Throughout the abstract screening process, the first and second author held weekly 1-hour meetings with the GAs. The focus of these meetings was to update the GAs on their overall progress toward meeting the deadline for reviewing abstracts and to discuss discrepancies from the previous week. Specifically, the first author sent the abstract coding file to the third author each week to calculate Kappa. Then, the first author selected three abstracts where there was disagreement. As a group, the first author, second author, and GAs read the abstracts and individually coded each abstract during the meeting. Then, we discussed whether or not the abstract should have been included or excluded, as well as reasons for our codes.

Finally, all abstracts were double-screened. Regardless of whether one graduate student or two GAs identified an article for inclusion in the full-text review process, a graduate student downloaded the full-text of the article and placed it into a group Zotero folder where all members of the research team could access the document. Prior to identifying the articles for full-text review, the first and second author reviewed the abstracts where only one graduate student identified an article for inclusion. If the first or second author could identify a reason for exclusion based on the abstract (e.g., one graduate student missed the fact that a study employed a single case design methodology and therefore deemed the study to be included), the study was removed from the full-text list. In total, the first and second author reviewed 143 abstracts and identified 37 for exclusion before full-text review.

Abstract Screening Results

Some studies were identified for exclusion at the abstract screening stage for more than one reason (therefore, the total number of reasons for exclusion is greater than 7,245). At this stage we excluded studies because they did not focus on academic learning or were generally irrelevant to early learning ($k = 711$), the abstract did not represent a *research* study ($k = 2{,}191$), conducting an intervention was not the purpose of the study ($k = 1{,}418$), intervention was the purpose of the study but the independent variable was not related to mathematics ($k = 1{,}726$), the study did not use a group design ($k = 867$), participants were not in preschool through 3rd grade ($k = 773$), and the study was a duplicate that was not immediately removed from the electronic search ($k = 75$).

Table 1.

Summary of Kappa Values for Interrater Agreement During Abstract Screening

Date	Number of Abstracts Kappa Refers To	Kappa Value	Total Abstracts Double-Coded
1/12/21	25	.27	25
1/18/21	150	.42	175
1/26/21	150	.88	325
2/3/21	600	.11	925
2/12/21	800	.69	1,725
2/15/21	1,000	.40	2,725
2/22/21	800	.71	3,252
2/27/21	1,000	.71	4,525
3/8/21	938	.59	5,463
3/22/21	1,048	.78	6,511
4/5/21	316	.71	6,827

Full Text Review and Training

The first and second author trained three GAs to complete the full-text review; the GAs were the same students who completed the abstract screening. Appendix B includes the full-text review protocol. The first and second authors developed the full-text review criteria and corresponding Excel database.

The first and second author made final determinations about the inclusion or exclusion of studies based on a series of information gathering (e.g., age of child participants, intervention setting, intervention agent, study design, outcome measures) about each of the full-text studies. In other words, the GAs did not make final determinations about the inclusion status of a full-text article; they coded each full-text article for information that would allow the first and second author to make this decision.

Prior to an initial 2-hour training on the full-text review process, the first and second author and GAs completed individual work in preparation of the meeting. Each person was assigned the following tasks: (a) review the full text criteria document; (b) review the Excel database; (c) read three studies (Niklas et al., 2016; Sonnenschein et al., 2016; Zippert et al., 2020); (d) code each of the three studies using the full-text review criteria and Excel database; and (e) take notes about suggested revisions to the criteria and Excel database.

The initial 2-hour training included the following topics: (a) a review of the purpose of a meta-analysis and specifically the purpose of the current project; (b) a discussion about what mathematics learning and teaching looks like in different environments; (c) a discussion of what an informal learning environment may look like; (d) a review of each of the full-text review criteria used to code the first three practice articles; (e) a discussion of challenges or necessary changes to the full-text review criteria or Excel database; and (f) the next assignments. During the initial 2-hour meeting, the research team discussed making changes to the full-text review criteria related to: how to determine if a study is focused on investigating the effectiveness of an intervention (compared to if a study is simply providing a description of an intervention without testing the effectiveness), providing detail related to clarifying between group assignment versus sampling procedures, and clarifying the necessary information GAs needed to provide related to the type of math content in the intervention. The first author made changes to the full-text review criteria and provided the new criteria and Excel database to the team, along with an assignment to code an additional two articles using the new materials (Berkowitz et al., 2015; Cheung & McBride, 2017). The research team reconvened the following day for an additional 1-hour follow up training session.

The purpose of the 1-hour follow-up session was to provide each member of the research team another opportunity to practice with the full-text review criteria and determine any additional challenges with using the criteria or Excel database. Each person independently coded the two studies and sent their codes to the first author before the meeting. The first author compiled the codes and identified three codes (of 22 possible codes) with discrepancies across raters. The 1-hour session focused on discussing the discrepancies and making changes to the coding sheet based on these discrepancies. These changes included: providing clarification around how to specify the grade level for studies (e.g., non-U.S. studies) that use nontraditional grade labels (e.g., "year 2 of kindergarten"); specifying the difference between studies that administered math measures and provided results versus studies that stated they administered a math measure but did not provide results and determining how to note the names of the measures given so that each GA was using a standardized approach. Again, the first author made changes to the full-text review criteria and the Excel database and sent the research team the updated materials. Then, the GAs were assigned an additional two studies to code independently (Dulay et al., 2019; Clements et al., 2011). After the GAs coded these studies, they met as a group (without the first or second author) to discuss any discrepancies. They also discussed any questions they had about the full-text review criteria and Excel database, as well as suggestions for final edits to the criteria. After the 1-hour meeting, a GA sent the first author remaining questions and suggestions.

The remaining studies were coded for full-text criteria by only two GAs (as opposed to all studies being coded by three GAs). Each graduate student coded between 30 and 40 full text articles each week. Throughout the full-text coding process, the first and second author held weekly 1-hour meetings with the GAs. The focus of these meetings was to discuss discrepancies from the previous week. At the end of each week, each GA sent the first author a list of the articles they were able to code for full-text. Then, the first author selected four to five studies where there was a disagreement about at least one code between GAs. During the meeting, the first author, second author, and GAs skimmed the article and individually coded each study for the code in question. Then, the group discussed the discrepancy and came to a consensus on the final code.

All remaining discrepancies that were not discussed as part of the larger group meetings were resolved between GAs in individual meetings. The first author merged all of the double-coded full-text codes across the GAs and indicated which codes had discrepancies and should be discussed by GAs. The GAs met as pairs to discuss the code and come to a consensus on the final code. Interrater agreement was 95.3% for full-text coding. The final codes were used by the first and second author to determine which studies would be considered for inclusion in the meta-analysis.

Identifying Studies for Inclusion: Full Text Review Results

We identified 373 studies for full text review (although, we were unable to obtain 9 articles; four articles were not available through interlibrary loan and five articles had abstracts in English but the full article was not available in English). Therefore, we only reviewed the full-texts of 364 studies. The first and second authors reviewed the final codes for the full-text coding to make determinations about the inclusion or exclusion of a study. First, the authors examined the three "stop coding" rules at the beginning of the full-text criteria. An answer of "no" to any of the following items resulted in a "stop coding" rule being employed that excluded studies from the meta-analysis: (1) Is the study about testing the effectiveness of an intervention or instructional program? (33 studies were excluded for this reason); (2) Are children in the study receiving some kind of mathematics content intervention, play, support, opportunities as a result of the study? (23 studies were excluded for this reason); and, (3) Did the study use a group design method? (28 studies were excluded for this reason). Second, the authors reviewed the codes for intervention setting and agent. If a study was identified as taking place in a school, classroom, child care center (e.g., Head Start), or in a university research lab setting, it was immediately excluded from consideration in the meta-analysis, as the purpose of this study was to examine the effects of mathematics interventions in informal learning environments (216 studies were excluded for this reason). Third, if a study only included children older than third grade, it was removed from inclusion (14 studies were excluded for this reason). Finally, 3 studies were excluded because they were duplicates of included studies, such as conference proposals that had eventually been published. From the full-text review phase, we excluded 326 studies.

In summary, we identified 47 studies in which children received a mathematics intervention in an informal setting (e.g., the home, museum) administered by a parent or other caregiver. However, only 22 studies were eligible for inclusion in the meta-analysis. Some studies (k = 25) were not eligible for inclusion in the *meta-analysis* for a few reasons. These studies were eliminated from the meta-analysis but were retained for follow-up manuscripts with different outcomes (e.g., math talk, literacy) or foci (e.g., a quality review). First, the study did not administer a mathematics achievement measure (k = 8; e.g., measures to test the effectiveness of the intervention were a proxy for math achievement such as rate of math talk, executive functioning, or were parent reports of the home numeracy environment). Second, the study was a school-based intervention with a home intervention component, but the effects of the home component could not be isolated from the effects of the school intervention ($k = 7$). Third, the study did not provide enough information to appropriately calculate effect sizes ($k = 7$; e.g., no posttest *M*, *SD*s even though the study was a pre to post-test design). When this occurred, the first author contacted the authors of the study in question for additional information. Although some authors replied with the necessary information, several authors did not reply and these studies were included in the quality review only. Finally, three studies had methodological issues that prevented us from including them in the meta-analysis.

- Austin (1988) - The dissertation did not utilize randomization or pre-test matching that would allow for causal inference. Clustering of students is a major concern; both in the small number of clusters (n = 4) and the lack of information needed to estimate the effect of clustering on variance estimates. Given the myriad of assumptions necessary to calculate an effect size we decided to exclude this study from analysis.
- Gervasoni & Perry (2017) - The authors present findings of group differences on a question-by-question basis. No single question represented a broader mathematics achievement construct that other effect sizes in our meta-analysis represent. Unable to aggregate item responses into a summative measure, we excluded this study from the meta-analysis.
- Lore et al. (2016) - Attrition in the treatment group was 76% and across groups was greater than 50%. Such levels of attrition make the validity of the effect size as a measure

of mathematics achievement gain highly questionable. Therefore, we removed the study from consideration.

Drafting the Coding Protocol

The first and second author drafted the coding protocol; both authors have extensive experience creating coding protocols for systematic reviews, meta-analyses, and qualitative studies. The final version of the coding protocol is located on an open access platform (https://scholarworks.boisestate.edu/sped_facpubs/141/). After an iterative process of feedback and revision between the first and second author, the coding protocol was sent to the third author (methodologist), four members of the advisory board, and three consultants - each with expertise in one or more of the following areas: early mathematics, parents and families, quality indicators, intervention, special education, and research synthesis. The first author addressed all feedback and refined the coding protocol. The coding protocol was further refined during the GA training (see process below).

The coding protocol comprises 5 sections, each of which is briefly explained below. See supplementary files for the full coding manual. 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/needed.

Excel Database

We used an Excel database for coding. Variables with a forced response (e.g., 0 = no; 1 = yes; 1 = random assignment; 2 = nonrandom assignment), were formatted in Excel to allow only codes included in the code book. The Excel database was also organized in separate tabs, as information from the sections below is reported on different levels (e.g., study level, outcome measure level, intervention group level). Each row of the Excel database across all tabs was linked with a unique study identifier.

Section 1: Basic Study Information and Participant Demographics

Phase 1 coding included a) basic study information such as authors, year, journal, type of publication, and location, and b) child participant sample and demographics information such as sample size, attrition, age or grade, gender, disability status, dual language learner status, race/ethnicity, socioeconomic status, and primary caregiver information.

Section 2: Methodological Information

Phase 2 included coding for basic methodological information such as study design, number of treatment and control groups, assignment procedures, treatment fidelity, quality of implementation, and interrater agreement information.

Section 3: Intervention Features

Phase 3 included coding for variables related to the implementation of the intervention, including: intervention agent, intervention location, intervention agent 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). During this phase, we also recorded information about the control condition, including if the control was described and if so what types of activities and content were implemented.

Section 4: Outcome Measures

Phase 4 included recording information about all outcome measures administered to measure the effectiveness of the intervention. Outcome measures other than mathematics achievement (e.g., working memory, literacy, home numeracy environment scales) were also recorded. For each outcome measure, we recorded the 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).

Section 5: Effect Sizes

Phase 5 included recording information that was necessary to calculate effect sizes and report on the quality of the study results, including: quality of data analysis, results reported effect sizes, results were reported in a clear fashion, outcome, design, sample size, and pre and posttest means and standard deviations. We also recorded information to calculate variance.

Study Quality

Within each phase of the coding protocol described above, we embedded and indicated which variables were associated with study quality. We used the Gersten et al. (2005) quality indicators to identify variables associated with study quality, and we also referred to PRISMA reporting guidelines to add any other variables or details related to study quality.

Full-Text Coding of Included Studies and Coder Training

After inclusion decisions were made and the codebook was drafted, the first and second author trained three GAs to complete coding of included studies. The GAs were the same students who screened abstracted and reviewed full texts, so they were familiar with the project goals and the related body of research. The training and coding processes were interactive, including a feedback loop to inform research team meetings, ensure on-going discussion, and refine the codebook and coding. To align with the five sections of the codebook (described above), our training was completed in five phases.

Training for each phase followed a similar structure and was completed in multiple sessions over a one-week span - one 2-hour session and two 1-hour sessions. Each phase of training included coding practice with the same 15 studies. The sessions were guided by a slideshow created by the first and second authors. This allowed the GAs to return to the trainings as needed throughout the coding process. Our slideshow also served the purpose of housing all information and resources that the GAs might need for training and coding in one place. Each training session was completed on Zoom; sessions were recorded as needed for later review. At the conclusion of each phase of training, the authors checked in with the GAs to identify what helped them learn and what they would like to do differently during the next round of training and coding. Below, we describe the initial 2-hour training sessions, the 1-hour follow-up trainings, and the coding process.

Initial Two-Hour Training Sessions

Each week of training began with a 2-hour session to introduce the sections and subsections of the codebook for that phase. Before each initial session, GAs were given tasks to prepare for the training, including: (a) read provided texts on relevant topics as needed; (b) review the full codebook, with attention to each subsection; (c) explore the corresponding Excel database, where they document their codes for each study; (d) read multiple studies (2-5) for practice; and (e) code each of the practice studies using the codebook and Excel database. As time permitted, the GAs discussed their preparation tasks as a small group before the team training.

As the GAs completed their training preparation tasks, they left comments on the codebook with questions about coding and suggestions for revisions to the Excel database (e.g., where descriptions were unclear, where code options were missing, etc.). They were instructed to carefully read through the table of contents and the notes in the codebook before they began reviewing the sections of the codebook, as well as come to training sessions prepared to lead a discussion of a subsection of the codebook. GAs also noted areas of the codebook for which they needed more information or further guidance.

When approaching the practice studies, GAs downloaded the studies from Zotero, where our team housed all data, to their own desktop folder. They read the entire study once, then went back through the study as they filled out the codes in their Excel database. As they read each study, they used one of the highlighting tools to indicate where they located the information needed to make each coding decision.

As part of preparation for initial trainings, GAs also read relevant texts as needed that were geared toward the content of the coding in that phase. For example, to prepare for Phase 3, which involved coding for math content and activities, GAs read and discussed the Common Core State Standards (CCSS) for Mathematics, the National Council of Teachers of Math (NCTM) Curriculum Focal Points, and a book chapter on developing early number concepts and number sense. A final training task included GAs bringing any terms that were unfamiliar from the texts read, the codebook or Excel database, or the studies reviewed to the meeting for further discussion.

The initial training sessions each began with a team member check in, a review of project purpose, and an overview of the project and the phase at hand. As needed, we spent time at the beginning of the initial sessions discussing concepts and ideas that were specific to the phase.

For example, Phase 2 required the GAs to code for "treatment fidelity," which the authors predicted would be a new concept for the GAs. Thus, the initial meeting for Phase 2 included a discussion of what treatment fidelity is, as well as examples and non-examples. Similarly, Phase 3 required identifying each informal math intervention group to code (i.e., many studies included more than one treatment group and not all treatment groups were a math treatment), so we discussed this process to ensure the GAs understood which group to code and why. During the review of content portion of the initial training, GAs also posed questions they brought from their preparation tasks about any new content or terms.

Discussing the codebook and the practice coding was the primary focus of the initial sessions. We first discussed each section and subsection of the codebook for the phase at hand. Some phases, such as Phase 1, involved discussing each individual code. Other phases, such as Phase 3, involved discussing groups of codes. For example, we discussed the math activities and math content overall, as opposed to discussing each individual activity and type of content. We felt this was unnecessary as the GAs had already completed readings on activities and content, as well as discussed these codes in a small group meeting. Our approach to reviewing the code book included each team member taking several pages to explain. During this explanation, we discussed any difficulties with coding in that section. We also brainstormed ways that the code book and associated Excel database could be revised for clearer description and more coherent organization.

After we reviewed the codebook for the phase at hand, we discussed the studies we practiced coding prior to the session. Most initial training sessions included a reflection on our practice coding (e.g., what was challenging and why). Next, the authors had GAs review their codes and reconsider them based on discrepancies that were identified by the authors across the three GAs. In other words, we identified codes with discrepancies and had the GAs consider the codebook and provide guidance to recode. We also had the GAs identify the page numbers where they found the information needed for coding, and we had them explain the codes they chose and why. This allowed for any misconceptions to be addressed. Based on our discussion of practice coding (and later coding), we either created a document with examples and explanations of "tricky" codes, added further explanation and examples to the codebook, or both. As we felt necessary, such as for Phase 2, we also provided the GAs with a study coded by the authors to use as a reference during coding. At the conclusion of each initial training session, the GAs were assigned 2-4 additional practice studies to code before the 1-hour session later in the week. They were instructed to first review any changes to the codebook and Excel database, made by the authors directly following each session.

Two One-Hour Follow-Up Training Sessions

During the same week as the initial training session for each phase, we also had two 1-hour training sessions. The first 1-hour session included all team members and had the purpose of discussing additional practice coding, as well as allowing the GAs to reflect on their coding after having discussed any coding challenges they faced. We began these sessions with a review of relevant coding notes based on discrepancies from the practice studies. All remaining time during the session was spent discussing discrepancies from the practice coding. The second author selected several studies with disagreement. As a group, the first author, second author, and GAs reevaluated their initial codes for the studies and recoded them during the meeting. Then, we discussed our coding, as well as reasoning for our codes. We also determined any additional revisions needed to the codebook or Excel database.

At the conclusion of the first 1-hour session, the GAs were assigned 2-4 additional practice studies to code before the second 1-hour session later in the week. They were also instructed to revisit all previous practice coding based on the two training sessions, and come to the next session ready to explain those codes. Before doing so, they were instructed to first review any changes to the codebook and Excel database, made by the authors directly following each session.

The second 1-hour session included only the GAs. Our rationale for this was to give the GAs time and space to discuss amongst themselves and process the learning from the week. Additionally, we found value in having the GAs explore challenging codes as a small group without the authors. The primary purpose of these sessions was determining final codes for all practice coding up to this point. GAs discussed the same 6-10 studies that were the focus of the previous training sessions; however, at this session, they discussed and deliberated on each discrepancy from each study. To conclude the training for each phase, the GAs coded additional practice studies, totaling 15 practice studies (the same studies for each phase).

Coding Process for Included Studies

Prior to the GAs moving forward after training, the discrepancies for each of the first 15 studies were discussed. Throughout the coding process, the second author sent the coding database to the third author to evaluate the level of agreement among the three GAs. The codes with low interrater reliability were discussed at a team meeting before the GAs continued coding.

Table 2. Description of Rounds of Coding for Each Phase

Round 1	Round 2	Round 3
● Triple coding of studies during/after training sessions ● Interrater reliability check to guide weekly team meetings ● Individual meetings to discuss discrepancies and identify final codes	● Double coding of studies ● Team meetings to discuss discrepancies and challenging codes ● Individual meetings to discuss discrepancies and identify final codes	● Double coding of studies ● Team meetings to discuss discrepancies and challenging codes ● Interrater reliability check to ensure quality before completing phase ● Individual meetings to discuss discrepancies and identify final codes

Throughout the coding process for each phase, the first and second author held weekly 1-hour meetings with the GAs. The purpose of these team meetings was to ensure that the research team was engaging in dialogue about the coding process a minimum of once per week. The focus of these team meetings was to discuss discrepancies from studies coded the previous week, as well as other specific codes that were challenging. Table 3 includes the three most challenging codes from Phases 1, 2, and 3 as indicated by the lowest interrater reliability in Round 1. The 12 codes with the most discrepancies in Round 1 each improved interrater reliability, as indicated by their interrater reliability in Round 3.

Table 3. Codes with Most Discrepancies from Each Phase of Coding

		Proportion Agreement of Pairs	
	Codes with Most Discrepancies	**Round 1**	**Round 3**
Phase 1 Study and Child Variables	Child DLL Reported	-0.10	0.69
	Child SES Reported	0.14	0.59
	Language in Home	0.44	0.90
Phase 2 Variables	Group Assignment	0.17	0.92
	Awareness Condition	0.61	0.79
	Nature of Instruction	0.73	0.83
Phase 3 Treatment Group Variables	Comparison: Sets or Objects	0.65	0.74
	Counting: Sequence or Verbal Counting	0.65	0.78
	Procedure Focus in Caregiver Training	0.70	.91
Phase 3 Control Group Variables	Nature of Instruction Reported	0.56	0.62
	Control Content	0.77	0.83
	Control Time	0.77	0.81

Another part of each team meeting was evaluating the codebook and the Excel database to determine if revisions were needed to make the coding process clearer. Based on GA questions from unique or challenging coding situations, the authors revised and added to the documents as needed. During coding, GAs left their questions directly on the codebook using the comments feature of Google Docs. This allowed the authors to respond to any uncertainties and revise the codebook promptly. Table 4 summarizes the revisions made to the codebook.

Table 4. Summary of Codebook Revisions for Each Phase

Coding Phase	**Revisions to the Codebook and Excel Database**
Phase 1: Basic Study Information and Participant Demographics	● Group level attrition was missing from the Excel database and was added. ● The correct number of options for child disability or at-risk were not provided in the Excel database as forced response options, so they were added. ● Pre and posttest attrition options were provided to reflect differences in how studies may report this information. ● Specification was made to use two decimal places consistently when reporting numbers. ● Specification was made to include the range of ages provided (minimum and maximum age) if age was broken down separately in the study by different child groups. ● Preprints were marked in the main Excel database for publication type, as the first and second author knew this information since the authors of the publications provided the manuscripts under review or accepted. ● Child SES was changed to family SES to reflect information about parents. ● Clarification was provided on how to add demographic information in the Excel database when authors provided information for the total sample and for the treatment and control separately.
Phase 2: Methodological Information	● Codes related to the number of caregivers in the household and who the primary caregiver was were removed. ● Clarification was added on the method of reporting the highest level of parent education. ● Asian American, Pacific Islander, and Native Hawaiian were combined as one group, given discrepancies in how authors may report this information. ● An anecdotal notes column was added for race/ethnicity. ● Clarification was provided on Head Start participation not always equaling low income (e.g., children may have access because they have a disability). ● The difference between Not Reported codes and Not Applicable codes was clarified.
Phase 3: Intervention Features	● Group names were provided in the main Excel database for consistency across GAs. ● Clarification was provided on “linking” groups across the tabs of the Excel database. ● Child and caregiver training/intervention code names were revised and further description was provided to more clearly indicate if information about the child, caregiver, or intervention in general should be coded. ● Further detail was provided on the definitions of duration, intervention length, intensity, and session length to differentiate between the terms. ● Guidance was provided for determining the appropriate use of “Not Reported” for duration codes. ● Clarification was provided on what "training" can entail. ● Specification was made on what should be captured in the “Other” activities and content codes.

After each round of coding, discrepancies were resolved between GAs in individual meetings. GAs submitted their Excel database once per week. The second author merged all of the double-coded files and indicated for the GAs which codes had discrepancies. Before discrepancies were discussed, GAs were instructed to first review the notes document created for by the authors to further explain any codes with consistent discrepancies and provide illustrative examples. As needed, the GAs also read through question and answer slides prepared by the authors based on the GAs' questions from the previous week. The goal of discrepancy meetings was to discuss the codes until a consensus was made on the final code decisions. GAs compared their codes and then reviewed the descriptions in the code book to make a final decision. In other words, the GAs did not make final determinations about any code individually. Instead, one of them documented their final decisions in a file based on their discussion. After each discrepancy meeting, the GAs emailed the first and second author with 1. a list of clarifying questions; 2. a database with final codes identified; and 3. notes on any codes they were unable to reach agreement on. The first and second author reviewed studies without agreement and made final coding decisions.

References

Cheung, S. K., & McBride-Chang, C. (2015). Evaluation of a parent training program for promoting Filipino young children's number sense with number card games. *Child Studies in Asia-Pacific Contexts*, *5*(1), 39-49.

Niklas, F., Cohrssen, C., & Tayler, C. (2016). Parents supporting learning: A non-intensive intervention supporting literacy and numeracy in the home learning environment. *International Journal of Early Years Education*, *24*(2), 121-142.

Polanin, J. R., Pigott, T. D., Espelage, D. L., & Grotpeter, J. K. (2019). Best practice guidelines for abstract screening large-evidence systematic reviews and meta-analyses. *Research Synthesis Methods*, *10*(3), 330-342.

Sonnenschein, S., Metzger, S. R., Dowling, R., Gay, B., & Simons, C. L. (2016). Extending an effective classroom-based math board game intervention to preschoolers' homes. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, *7*(2), 1.

Therrien, W. J., Cook, B. G., & Cook, L. (2020). Utilizing meta-analyses to guide practice: A primer. *Learning Disabilities Research & Practice*, *35*(3), 111-117.

Zippert, E. L., Douglas, A. A., Smith, M. R., & Rittle-Johnson, B. (2020). Preschoolers' broad mathematics experiences with parents during play. *Journal of Experimental Child Psychology*, *192*, 104757.

Appendix A

Abstract Screening Protocol

Abstract Screening Criteria

1. Is the study **academic learning-focused**?

Yes	No	Maybe
The study is related to academic learning (i.e., related to learning of students in grades PreK-12th grade; related to academic content areas such as science, reading, math; academic achievement;).	The study is not related to academic learning and is generally irrelevant for the purpose of this meta-analysis (e.g., related to students in PreK-12th grade but not to other areas such as behavior, physical education, music or art education, overall health and well-being; healthcare, medicine, the military, agriculture).	It is not clear from the abstract what the study is related to.

2. Is the article a **research study**?

Yes	No	Maybe
Yes, the study is a research study. Researchers investigated some sort of variable.	No, the study is not a research study. This may include: booklets, descriptions of resources or teaching guides, policy briefs, state reports, brochures, commentaries, and book or test reviews. Other indicators of this: (a) no mention of any participants as research subjects; (b) no mention of any research design.	It is not clear from the abstract if the study is a research study.

3. Is the study about **testing the effectiveness of an intervention** or instructional program? I

Yes	No	Maybe
The study is about testing the effectiveness of an intervention or instructional program. The program can be focused on adults who then interact with children. For example, a parent training workshop on effective math activities to do at home. Intervention terms may include: teaching experiment, tutoring, supplemental supports or instruction, trial, training, workshop.	No, the study is not about an intervention or instruction program (e.g., commentaries, achievement generally, longitudinal achievement, assessment development, teacher attitude, student perceptions). Or, the purpose of the article was not to report effects of an intervention, which may (but not always be an indicator) include: briefs, brochures, presentations, resource, review.	It is not clear from the abstract what the study is related to.

4. **What is the independent variable?**

If you selected "yes" above that a study is an intervention study, it should be clear from the abstract what the independent variable is. In other words, what were the researchers manipulating?

a. Independent variable = the intervention; what is the change from typical instruction that a researcher/teacher/parent is manipulating to see a change in the dependent variable.
b. Dependent variable = what is being measured to determine a change in something (usually achievement)

5. Is the dependent variable in the study a **mathematics content intervention**? This focus is on the independent variable (what is being manipulated by the researcher); this focus is not on the dependent variable (outcome). In other words, a self-regulation intervention (independent variable) with outcomes as measured by math performance is not a math intervention. It is still a self-regulation intervention.

Yes	No	Maybe
Participants received an intervention related to mathematics. mathematics (e.g., counting, comparison, mathematics vocabulary, math book reading, math talk, numbers, operations, spatial reasoning, data analysis and probability, geometry, measurement, shapes, early numeracy, composing/decomposing, addition, subtraction, multiplication, division, fractions, place value, number lines, algebraic thinking, problem solving, story problems, etc.). Please note that mathematics interventions may be paired with another type of intervention such as game-based counting intervention, math story book intervention, etc.	The independent variable in the intervention was something other than mathematics content, including (but not limited to): quality of early childcare, reading comprehension, fluency or decoding, phonological awareness, social studies, behavior interventions, self-regulation, self-monitoring, social skills, communication, physical or occupational therapy, speech/language, mental health, or cognitive interventions (e.g., working memory).	Students were receiving an intervention, but is not clear from the abstract what type of intervention this was or what the ind. variable was.

6. Did the study use a **group design method**?

Yes	No	Maybe
The study used a group design, experimental design, or quasi-experimental design (indicators of this include terminology such as: QED, random cluster trial, RCT, mixed methods, treatment and control group, random assignment, pre to post-test design, post-test only, ANOVA).	No design is referenced. If you can't identify any study design terms, it is likely not a research study. Or, the study used a non-group design method. This might include a qualitative study, a case study, literature reviews, systematic reviews, meta-analyses, or single case design method (also referred to as: multiple baseline, multiple probe, reversal design, AB, ABAB, alternating treatment, visual analysis, theoretical pieces). Very small *N*s (such as less than 6) are typically a good indicator of a single case design study. Other non-intervention group designs include: latent growth curve modeling, growth curve analysis, longitudinal, path analysis.	A design is referenced but it is not clear if it is group or other; terms may include: action research, research, observations, or surveys.

7. Are there participants in the study and are *some* of the participants who received the intervention in **preschool through third grade**?

Please note that it will be difficult to determine if all participants meet this requirement just from the abstract, which is why the word *some* is emphasized.

Yes	No	Maybe
There are participants in the research study and at least some of the participants in the study were between the ages of 3 years, 0 months and the end of third grade (approximately 9 years, 0 months).	There are no participants in the study. If no participants are referenced, it is likely not a research study. Participants in the study were only infants or toddlers under the age of 3 or children older than third grade (e.g., if the abstract states "participants were 4th and 5th graders") or adults (e.g., if the abstract states "college students").	Participants are referenced, but it is unclear if the study includes some participants between the ages of 3 years and the end of 3rd grade, or if the abstract/title does not reference age or grade of participants at all.

8. Is this study original/have you reviewed it previously?
   a. Yes, this study is original and I haven't reviewed it previously.
   b. No, this study is a **duplicate**. – Make a note that says DUPLICATE.

Appendix B
Full Text Review Protocol

Overview
You will recode each full text article as if it is the first time you are seeing the abstract/study. In other words, your codes for the abstracts will not be considered when you code for the full text review. Some of the criteria are the same as the abstract review process, but most criteria are new. There are also three "stop coding" notes, at the very beginning of the full-text review; otherwise, **you will fill out all columns for each study, even if you can immediately see another reason for exclusion** (e.g., there isn't a math outcome measure).

Keywords to search for in the PDF: random, intervention, delayed, post.

Full Text Review Criteria

1. Is the study about **testing the effectiveness of an intervention** or instructional program? (Note: if this is about testing the effectiveness of an intervention, the research questions should mention examining the effectiveness of an intervention or student improvement in a dependent variable as the result of an intervention).
    1. Yes, the study is about testing the effectiveness of an intervention or instructional program.
    2. No, the study is not about an intervention or instruction program (e.g., commentaries, achievement, longitudinal achievement, assessment development, teacher attitude, student perceptions).
        - IF YOU SELECT NO, stop coding the article.
2. Are *children* in the study receiving some kind of **mathematics content intervention**, play, support, opportunities as a result of the study?
    1. Yes, as a result of the study child participants were receiving an intervention or extra support related to mathematics.
        - Please note that mathematics interventions may be paired with another type of intervention such as game-based counting intervention, math story book intervention, an iPad math fact fluency app, etc.
        - Please note, that parents or other adults may *also* have been receiving an intervention (e.g., training) to inform them about how to provide extra math support to children. However, to receive a code of yes in this column, children must have received some sort of math support.
    2. No, child participants in the study were not receiving any intervention at all.
    3. No, child participants in the study were only receiving another type of intervention such as reading comprehension, behavior interventions, self-regulation, self-monitoring, social skills, communication, physical or occupational therapy, speech/language, mental health, or cognitive interventions (e.g., working memory).
        - IF YOU SELECT NO, stop coding the article.
3. Did the study use a **group design method**?
    1. Yes, the study used a group design, experimental design, or quasi-experimental design (indicators of this include terminology such as "treatment and control

group" "random assignment" "pre to post-test design"). This also includes instances where there is just one treatment group and no control group (i.e., all children in the study received the intervention).
2. No, the study used a non-group design method. This might include a qualitative study, a case study, literature reviews, systematic reviews, meta-analyses, or single case design method (also referred to as: multiple baseline, multiple probe, reversal design, ABAB, alternating treatments, adapted alternating treatments).
    - IF YOU SELECT NO, stop coding the article.

4. What is the **data collection procedure** (Design Information)?
    1. Pre to post-test
    2. Post-test only
    3. Pre to post-test with delayed post-test (might be called: maintenance, follow-up, T1, T2, T3)
    4. Pre-test only
    5. Post-test only with a delayed follow up post-test
    6. Other
    7. Not reported
5. What **group type** is used?
    1. Treatment(s) and control(s)
    2. Treatment group only
    3. Other
    4. Not reported
6. What was the **group assignment** (i.e. This refers to how groups were assigned to the treatment and control conditions, not how participants were selected for the study sample)?
    1. Random assignment (includes statements such as "random assignment blocking on classroom")
    2. Nonrandom assignment
    3. Self-selection into the treatment
    4. Matching, or yoking, yoked pairs, propensity score matching
    5. Regression discontinuity
    6. Other
    7. Not reported
7. If "Other" was chosen for the previous code, what was the **other group assignment** to the treatment?
    1. Make a statement regarding how the authors described the assignment.
        - Fill out this column if you selected "Other" for the previous code
        - If "Other" was not selected for the previous code, put "NA" in this column
8. What are the **intervention activities?**

Provide some information, less detail is okay, about what the intervention looked like. For example, was it an intervention package ("Building Blocks") or activities such as card games, number games, story books, etc. Do not worry about reporting the specific math content that is included in the intervention unless no other information is given.

1. Note that at this time, we are including what is called "logico-mathematics" which refers more to things like classification, seriation, and sorting. It is

commonly referred to as "logico-mathematics" and typically an article will also refer to Piaget's theories of development.

9. What is the **intervention type**?
    1. Make a statement or two about the type of intervention that was being tested and what the intervention agent was expected to do. See example statements:
        - Parents attended a workshop on how to embed math activities in cooking routines at home. Parents reported how often they talked about math at home.
        - Parents received coaching at a school event or museum about how to use math talk during book reading. Parents video-taped book reading at home with their children.
        - Children were sent home from school with a math kit; the kit contained instructions for parents about how to play a board game or use cards. Parents and children played board games at home.
10. What was the **length of the intervention** or how often were adults instructed to use the intervention with children?
    1. Make a statement on how long the child math intervention lasted. Consider the number of sessions and the amount of time each session lasted.
        - If parents are provided with a recommendation, provide that as opposed to the results. For example, the article might state "Parents were told to implement the intervention for 15 sessions over 3 weeks" and in the same article, the authors might report, "On average parents implemented 13.2 sessions over 3 weeks." **The information that we want at this point is the recommendation.**
        - At this time, we are not considering how long parents received training.
        - Include a number (reported as a numeral) and a unit (e.g., 1 session; 3 weeks, 2 months) and if reported also include session length (e.g., 1 session lasting 45 min; 3 weeks of 3 sessions per week, each session was 20 min)
        - If there is a pilot study included before the intervention study, only report the intervention period.
        - Include phrase "not reported" if you are unable to locate this information
11. Are *some* of the participants who received the intervention in **preschool through third grade**?
    1. Yes, at least some of the participants in the study were between the ages of 3 years, 0 months and the end of third grade (approximately 9 years, 0 months). This includes instances without specific ages (e.g., 5 years old) but use of grade level such as "preschool" or "kindergarten" to specify age.
    2. No, participants in the study were only infants or toddlers under the age of 3.
    3. No, participants in the study were only children who were older than third grade (e.g., if the abstract states "participants were 4^{th} and 5^{th} graders") or adults (e.g., if the abstract states "college students").
    4. No, participants in the study were only infants or toddlers under the age of 3 and children who were older than third grade.

5. Not reported; this may refer to articles with no information about child participants' age or ambiguous identifiers such as "young children."

12. What is the **specific <u>grade level or range of grade levels</u>** for child participants?
    1. In this column, specify the grade or range of grades in the study (children only; do not make conversations to grade based on age; report only as authors give you information; e.g., kindergarten in the US is a different age than in the Netherlands)
        - Use NA if you did not select Yes for the code above.
        - Use "not reported" if you are unable to locate this information.
        - Use PreK for any grade before kindergarten, even if other terms (e.g., preschool, Head Start, early learning center, etc.) are used.
13. What is the **specific <u>age or range of ages</u>** for child participants?
    1. In this column, specify the age or range of ages in this study (children only; report as authors report the information and do not make conversions; e.g., 57-62 months, 4.5 - 6 years)
        - Use NA if you did not select Yes for the code above.
        - Use "not reported" if you are unable to locate this information.
14. What is the **<u>average age of children</u>** in the study (specify years or months as it is reported in the study and do not make conversions to another unit)
    1. In this column, specify the average age in the study (children only)
        - Use NA if you did not select Yes for the code above.
        - Use "not reported" if you are unable to locate this information.
15. What is the **<u>study setting</u>**?
    This code is specific to where the children eventually received the math help. For example, a study might describe an instance where parents attend a workshop at their child's school; but then the study specifies that parents were instructed to do activities *at home* with their children. ***This code is specific to where children eventually get the math content delivered to them.***
    1. School classroom (including center-based childcare, Head Start, preschool, elementary schools) during the child's school day.
    2. After-school program at a school or public space.
    3. Child's home.
    4. Home-based daycare setting, including after school childcare in a home-based daycare setting (different from a childcare center).
    5. Public space such as a grocery store, library, garden, zoo, museum, or community center.
    6. Other
    7. Not reported
16. If "Other" was chosen for the previous code, what was the other **<u>setting?</u>**
    1. Make a statement about where the child received the extra math support/intervention.
        - Fill out this column if you selected "Other" for the previous code
        - If "Other" was not selected for the previous code, put "NA" in this column

17. Who was the **intervention agent**?
   This refers to who was administering the math content to the child. We are not interested at this point who trained the parents, childcare providers, museum staff, etc. We are interested in **who facilitated the math learning** for the child.
   1. Classroom teacher, paraprofessional, other school staff
   2. Researcher, graduate student, research assistant
   3. Parent, guardian, grandparent, or other adult family member
   4. Older sibling
   5. Home-based daycare provider
   6. Staff at a public space (librarian, zoo staff, museum staff)
   7. After school staff or volunteers
   8. Other or Mix of types of intervention agent
   9. Not reported
18. If "Other" was chosen for the previous code, who was the **other intervention agent**?
   1. Make a statement about who facilitated the math learning.
      - Fill out this column if you selected "Other" for the previous code
      - If "Other" was not selected for the previous code, put "NA" in this column
19. Were **child math achievement outcome measures** collected and reported?
   The authors administered at least one measure of mathematics achievement to measure the effectiveness of the intervention. This may include, but is not limited to: counting skill, patterning recognition, multiplication, fact fluency, basic addition and subtraction, word problem solving, etc. **For this code, we are only interested in child level measures that relate to math achievement.** We are not considering "math talk" here.
   1. Yes and yes, the authors collected and reported at least one child level math outcome.
   2. Yes and no, the authors collected math data but did NOT report results.
   3. No, the authors did not collect or report child level data on at least one math outcome measure.
   4. Maybe, it is difficult to tell from the study if the measure was a math measure or another type of measure, or if data were reported.
20. Were **child literacy achievement outcome measures** collected?
   The authors administered at least one measure of literacy achievement to measure the effectiveness of the intervention. For this code, we are only interested in child level measures, such as but is not limited to: phonemic awareness, letter knowledge, spelling knowledge, phonological awareness, comprehension, accuracy/word identification, rate/WPM, reading strategy use, vocabulary knowledge, oral language, writing skill, syntactic knowledge, etc.
   1. Yes and yes, the authors collected child level outcome data on at least one literacy outcome measure.
   2. Yes and no, the authors collected literacy data but did NOT report results.
   3. No, the authors did not collect or report child level data on at least one literacy outcome measure.
   4. Maybe, it is difficult to tell from the study if the measure was a literacy measure or another type of measure, or if results were reported.

21. Were **<u>child "math talk" outcomes</u>** considered?
    1. Yes and yes, the authors collected and reported child level data on "math talk" such as frequency of math words in a video-taped session.
    2. Yes and no, the authors collected "math talk" data but did NOT report results.
    3. No, the authors did not collect child level data on "math talk."
    4. Maybe, it is difficult to tell from the study if "math talk" data were collected or reported.
22. What were the specific **<u>child measures administered/collected?</u>**
    1. List all of the child measures administered and collected separated with a semicolon (;) Report this as the APA headings, as the measures are titled by the authors of the study.
    2. Note: for standardized norm-referenced tests you can simply include the test acronym as opposed to writing out the full test name and subtests. Examples include: WJ-II (Woodcock Johnson Tests of ability and achievement); WISC (Wechsler Intelligence Scale for Children); TEMA (Test of Early Math Ability); WIAT (Wechsler Individual Achievement Test); KeyMath; SAT-10 (Stanford Achievement Test).

For tests that are researcher developed or less commonly used norm-referenced tests, list the names of the test (such as with the heading provided in the study).