Check for updates

Article

Hammill Institute on Disabilities

Learning Disability Quarterly
1–14
© Hammill Institute on Disabilities 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/07319487261488012
journals.sagepub.com/home/ldq
Sage

Advancing Early Numeracy Research: A Systematic Review of Risk Identification and Intervention Features

Gena Nelson, PhD[1], Soyoung Park, PhD[2], Sarah Quinn, PhD[3], Patrick Ehrman, PhD[4], Taylor Lesner, PhD[1], Marah Sutherland, PhD[1], Kevallyn R. Drake, MIT[1], Peter Boedeker, PhD[5], Lina Shanley, PhD[1], and Ben Clarke, PhD[1]

Abstract

Consistent reporting of participant characteristics and criteria used to screen for academic risk is essential for interpreting and applying intervention research findings. Clarifying how risk is defined helps determine which interventions are effective—both for whom and under what conditions. This review examined how academic risk is defined and operationalized in PreK–1 early numeracy intervention research and explored its relationship with intervention features, instructional strategies, and mathematics content. A total of 112 studies published between 2006 and 2025, including more than 21,000 participants, were analyzed. Results showed substantial variability in screening criteria. Many studies (40.6%) did not specify how academic risk or need for intervention was determined. Among studies reporting risk levels, 22% of samples were high risk, 38% were moderate, and 6% were very low risk in mathematics, while 17% had socioeconomic status (SES)-based risk. Variation in how risk related to intervention characteristics highlights the need for clearer reporting and more consistent risk identification practices.

Keywords

at risk, content-area instruction, instructional strategies, learning disabilities, identification

Introduction

Early numeracy is one of the strongest predictors of later mathematics achievement and is linked to important long-term outcomes, including advanced high school mathematics course-taking and college enrollment (Davis-Kean et al., 2022). Yet, many students do not develop these foundational skills early on, and the nation continues to experience underachievement in mathematics, as demonstrated by recent results from both national and international assessment data. Achievement gaps in mathematics exist for some children as early as kindergarten (Burchinal et al., 2011), and in the absence of intervention, these gaps continue to widen (Nelson & Powell, 2018). These achievement gaps are bleaker for students with academic risk factors, including having a disability or being from a low socioeconomic status (SES) family (National Center for Education Statistics [NCES] & US Department of Education, 2025). In fact, by fourth grade, 35% of students from low-SES backgrounds performed below "Basic" on the 2024 National Assessment of Educational Progress in mathematics, whereas only 12% of their higher SES fourth-grade peers performed below the "Basic" level (NCES & U.S. Department of Education, 2025). Moreover, in a large longitudinal study, authors reported that students from low SES backgrounds or those with disabilities were at a heightened risk for persistently low mathematics achievement (i.e., <25th percentile) through eighth grade (Morgan et al., 2016).

Students with persistent difficulty in mathematics include those with mathematics learning disabilities (MLD), as well as students who struggle with low mathematics achievement without a formal disability diagnosis (i.e., mathematics difficulties [MD]). This distinction between

[1]University of Oregon, Eugene, OR, USA
[2]University of Central Florida, Orlando, FL, USA
[3]Eastern Michigan University, Ypsilanti, MI, USA
[4]Purdue University, West Lafayette, IN, USA
[5]Baylor College of Medicine, Houston, TX, USA

Corresponding Author:
Gena Nelson, Center on Teaching and Learning, College of Education, University of Oregon, 5292 University of Oregon, Eugene, OR 97403, USA.
Email: gsnelson@uoregon.edu

the early identification of MD and formal diagnosis is especially important in early grades, such as prekindergarten through first grade (PreK–1). Not all students who are identified as having MD in primary grades meet the criteria for MLD later on (Mazzocco & Myers, 2003), suggesting that trajectories can be altered for students with various forms of academic risk. A growing body of research suggests that early intervention can significantly improve student outcomes (Doabler et al., 2021), particularly when methods of support are aligned with students' level of risk (Dennis, 2015). Considering the significant cost associated with providing early numeracy intervention (Clarke et al., 2020) and the results of previous research indicating smaller treatment effects of interventions over time (Dietrichson et al., 2021), intervention in the early grades is essential to mitigating chronic MLD.

A key component of providing early intervention in schools and research is to screen and identify the students who have academic risk and may require supplemental support. Yet, academic risk is not universally defined. For example, researchers apply different criteria to determine students' SES and the level of SES that indicates risk (Tan, 2024), with states and experts varying in their perspectives about how learning disabilities should be diagnosed (Benson et al., 2026). Research on MLD and MD identification remains in its early stages and is also widely debated (Lewis & Fisher, 2016; Nelson & Powell, 2018). These challenges are particularly important, given the need for accurate and reliable identification to support timely and appropriate intervention for students who need additional instructional support.

Although prior research has examined MLD and mathematics difficulty identification criteria (Lewis & Fisher, 2016), the current study examines academic risk screening criteria in the context of early numeracy interventions. A review of this nature can provide insight into how academic risk is operationalized and, subsequently, shed light on the areas of early numeracy that remain underexplored for students with different types of academic risk. To meet these aims, we synthesized screening criteria and intervention features from 112 early numeracy interventions from a preregistered meta-analysis (Nelson, Boedeker, 2026; Nelson, Boedeker, Lesner, Clarke, Sutherland, et al. , 2026).

Intervention Effects by Risk Status in Early Numeracy Intervention

A growing body of intervention research has examined the effects of early numeracy interventions for students at varying levels of academic risk, including students from low SES and those identified with MD. Results have been inconsistent. The results of previous studies have suggested that intervention effects may differ based on how students' risk status is determined. For example, Toll and Van Luit (2012) found that an intervention for kindergarteners who fell between the 25th and 50th percentiles resulted in significantly higher numeracy test scores, but not among students below the 25th percentile. In contrast, Cary et al. (2021) reported greater treatment effects for students at the highest and lowest levels of risk as compared with those at moderate levels of risk in their kindergarten technology-based intervention. Two recent meta-analyses focused on early numeracy interventions (Benson, Farmer, et al., 2026; Nelson & McMaster, 2019) reported that academic risk status, including low SES and MD, accounted for a significant proportion of variance in effect sizes. However, Pereira et al. (2026) also conducted a meta-analysis of 87 early numeracy interventions and reported that intervention effects were not moderated by children's MD or SES risk status. In a meta-analysis of 191 PreK–12 mathematics interventions, Williams et al. (2022) also reported that student SES did not account for a significant proportion of variance in effect sizes ($p = .89$). These inconsistent findings suggest the need for a better understanding of how students are identified as being at academic risk in early numeracy intervention research.

Previous Systematic Reviews of Mathematics Risk Identification

In early numeracy research studies, students' academic risk status is determined using various screening approaches—from school-based identification methods (e.g., cutoff scores on a screening measure) to general indicators of risk such as low SES. Such variation has important implications for interpreting intervention effects, as students may be classified as being at a different level of risk based on using different screening criteria. From prekindergarten to first grade, the focus is on identifying students who are at risk of later being identified with MLD. Thus, how academic risk in mathematics is identified is critical, as it directly informs the need for intervention in school-based studies and practice. Prior reviews have documented considerable variability in how mathematics risk is identified (Lewis & Fisher, 2016; Nelson & Powell, 2018), including differences in classification criteria.

Lewis and Fisher (2016) conducted a systematic review of 164 studies on MLD published between 1974 and 2013 and found great variability in classification criteria. Specifically, they found that authors used a range of cutoff scores from the 2nd to the 46th percentiles, with the 25th percentile as the most common percentile researchers used to identify MLD. Shortly after, Nelson and Powell (2018) reviewed 35 longitudinal studies to understand the growth of students with and without MD on mathematics measures and examine how MD was identified. They found that the identification of MD was a strong predictor of later mathematics performance and observed a lack of consistency in the cutoff criteria used for classification across studies, ranging from the 10th to the 35th percentile, with the 25th percentile being the most common.

Purpose and Research Questions

Although previous reviews have advanced our understanding of mathematics risk identification, no systematic review has examined how students are screened and identified for academic risk, specifically within PreK–1 early numeracy intervention research. A synthesis of the screening criteria used in early numeracy intervention research is needed to better understand how participants are identified across intervention studies. The present study also extends previous reviews by synthesizing key intervention features across different screening criteria. Collectively, investigating and reporting on these aspects of early numeracy interventions may enhance future researchers' reporting related to operationally defining academic risk and identify underdeveloped areas of early numeracy intervention research for various populations. The research questions guiding this systematic review are:

Research Question 1 (RQ1): Which screening criteria do researchers use to identify PreK–1 students for early numeracy intervention studies, and what is the frequency with which these criteria are used?

Research Question 2 (RQ2): What patterns and gaps exist in the key features of PreK–1 early numeracy interventions (i.e., grade, malleable intervention features, instructional strategies, and mathematics content) across screening criteria?

Method

This systematic review used an existing dataset from a previous PreK–1 early numeracy intervention meta-analysis that was registered on *Open Science Framework* (Benson, Farmer, et al., 2026). Thus, this systematic review retained the inclusion criteria, literature search procedures, and article coding procedures used by Benson, Farmer, et al. (2026) for consistency. These aspects of the pre-registered meta-analysis are briefly described below.

Inclusion Criteria

1. The study included participants from PreK to Grade 1.
2. The study reported on the effectiveness of an intervention program—defined as either (a) a supplement to or substitute for core instruction or (b) a change to typical core instruction such as increased instructional intensity, dosage, or smaller group size (Wanzek & Vaughn, 2007)—including at least four sessions across at least 2 weeks. Tier 1 programs were excluded.
3. The intervention targeted whole number concepts in the number, relations, or operations domains (National Council of Teachers of Mathematics [NCTM], 2006; National Research Council, 2009). Figure 1 includes a list of skills we coded for across the domains.
4. The study occurred in a school-based setting (e.g., classroom and intervention setting).
5. The study used an experimental or quasi-experimental group design with at least one treatment and one control group. Thus, studies that used nonrandom assignment and a post-test-only design were excluded (e.g., Lysenko et al., 2016), as were those that used single-case designs (SCD; e.g., Yazıcıoğlu & Akdal, 2023).
6. The study included a control condition that (a) was either a business-as-usual condition or an active control that was not an intervention but met inclusion criteria, and (b) was similar to the intervention group in academic risk; studies that only compared at-risk groups to not-at-risk groups were excluded (e.g., Bryant et al., 2008).
7. The study reported appropriate information to calculate effect size (ES; e.g., sample sizes, means, *SD*s, *F* statistics, and *t*-tests) for at least one mathematics outcome.
8. The study was reported in English.

Literature Search Procedures

We conducted the literature search in August 2024 by examining peer-reviewed and gray literature (e.g., conference proposals and dissertations). We included literature published in 2006 or later to align with the release of the NCTM (2006) *Curriculum Focal Points*, which identified key mathematics concepts designed to support PreK–1. The literature search included electronic library databases, a reference list review of previously published systematic reviews, a forward citation search, and posts on relevant listservs for in-progress manuscripts.

First, we searched electronic databases, including Academic Search Premier, Educational Resources Information Center, American Psychological Association (APA) PsycNET, Web of Science, and ProQuest Dissertations and Theses Global. We used the Boolean string: *Intervention* OR *training* OR *instruction* OR *tutoring* OR *program* OR *remedial* OR *supplemental* OR "*tier 2*"OR "*tier 3*" AND *math** OR "*early math**" OR *numeracy* OR "*number sense*" OR *numerical* OR "*whole number*" AND *preschool** OR "*early childhood*" OR *prekindergarten* OR *kindergarten* OR "*first grade*" OR "*Grade 1*" OR "*prek*" OR "*primary school.*"

Second, we reviewed the reference lists of nine relevant mathematics intervention reviews (Charitaki et al., 2021; Filiz & Güneş, 2023; Jitendra et al., 2021; MacDonald & Murphy, 2021; Monei & Pedro, 2017; Nelson & McMaster, 2019; Schnepel & Aunio, 2022; Svane et al., 2023; Williams

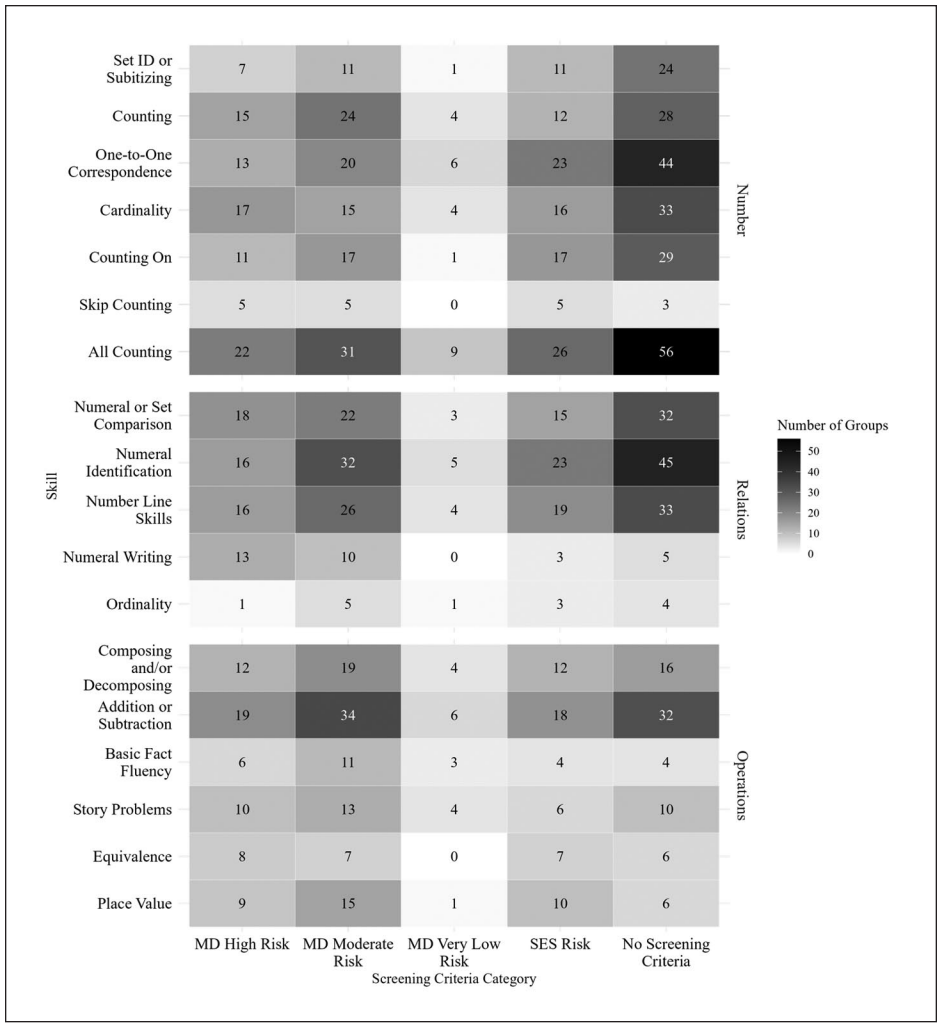


Figure 1. Heatmap of targeted content areas by screening criteria.
Note. The number within each cell represents the *n* of intervention groups receiving instruction in the indicated skill area at the indicated screening criteria group.

et al., 2022). We created a list of all included studies across the reviews. Third, we used Web of Science to create a list of studies that had cited Nelson and McMaster (2019), a meta-analysis on which the current project builds. Finally, we posted messages on three organizational listservs that allowed open messages to researchers. We asked members to send any in-progress, under-review, or recently accepted manuscripts that met our inclusion criteria. Across these other literature search methods, we reviewed any newly found studies for inclusion.

Literature Search Results

We included the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Page et al., 2021) diagram from the meta-analysis (Benson, Farmer, 2026) as Supplemental Figure S1. Ten researchers were trained to conduct the abstract screening and full-text review using detailed protocols aligned with the inclusion and exclusion criteria. The protocols are available on the project's Open Science Framework page (Nelson et al., 2025; https://osf.io/4m5gn/overview). The training included virtual synchronous team meetings, group practice, and independent practice. During the screening and full-text review processes, the researchers also met weekly to discuss and resolve all discrepancies. In total, we reviewed 15,252 abstracts. The majority of studies were excluded at the abstract screening stage (k = 14,407; 94.5%). All abstracts were reviewed independently by two coders, and the agreement on inclusion and exclusion at this stage was 96.5%. Any abstract with a discrepancy was discussed between the two coders to make a final decision about the study; agreement was 100% after discrepancy meetings.

From abstract screening, we identified 817 articles for full-text review. These 817 articles included 886 studies

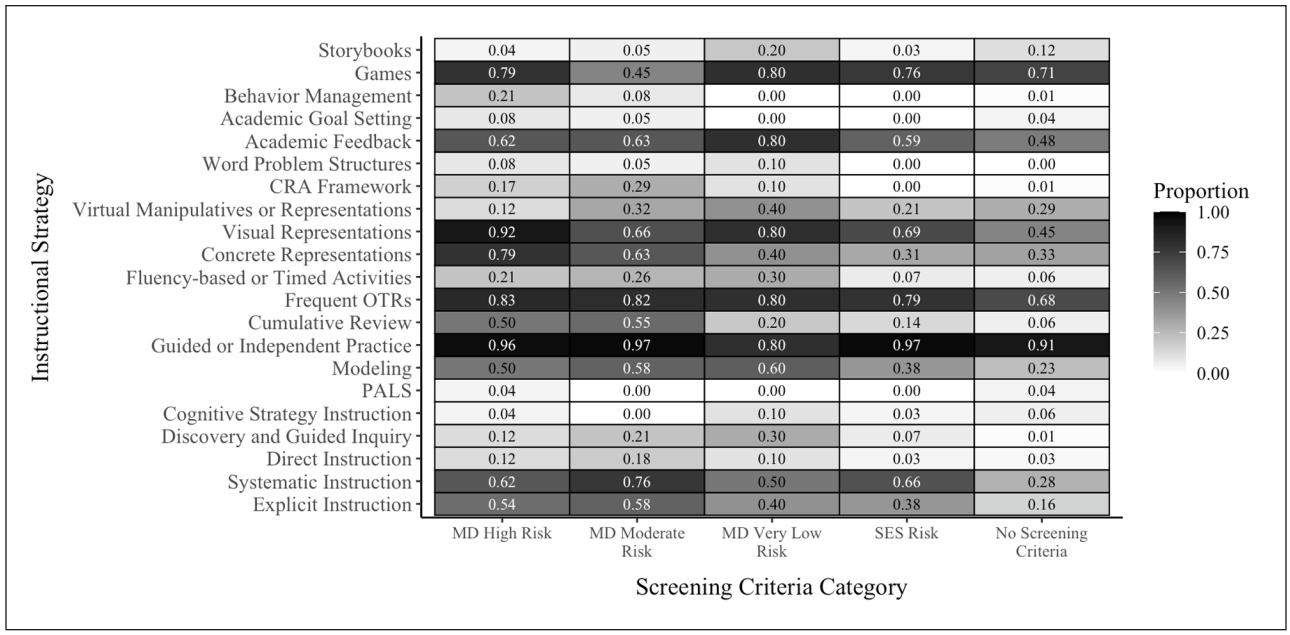


Figure 2. Instructional strategy use by screening criteria.
Note. The x-axis represents screening criteria groups, and the y-axis represents instructional strategies. MD = mathematics difficulties; PALS = peer-assisted learning strategies; OTRs = opportunities to respond; CRA = concrete-representational-abstract

(e.g., Study 1 and Study 2 were published in the same article). All full texts were reviewed independently by two researchers; agreement for inclusion was 87% and rose to 100% after discrepancy meetings. Figure 1 provides details about reasons for study exclusion during full-text review. From the full-text review process, we identified 111 articles (with 112 total studies) for inclusion in the Author et al. (2026b) meta-analysis. All studies from the previous meta-analyses were included in the current systematic review.

Article Coding

The studies in this systematic review were coded for approximately 165 variables for the purpose of the larger meta-analysis project. The coding manual, which includes variables, code response options, and detailed descriptions of codes, can be accessed online (Nelson et al., 2025; https://osf.io/4m5gn/overview). For the current analysis, we extracted the following variables from the dataset.

Basic Study Information and Sample Demographics. For each study, we recorded the publication year, publication type, study location, and sample demographics (i.e., age/grade, gender, race/ethnicity, multilingual learner status, SES, disability/eligibility for special education).

Grade Level. The grade level of participants from studies conducted in the United States was extracted as it was reported by study authors. For studies conducted outside the United States, we converted information to U.S. terminology for comparison where necessary. In the United States, many children attend preschool at age 4, enter kindergarten at age 5, turn 6 years old during the school year, and then begin Grade 1 at age 6, turning 7 years old during the school year. In some countries, the terms "preschool" and "kindergarten" are used differently from how these terms are used in U.S. studies. For example, in Hong Kong and France, there are 3 years of kindergarten, but only the third year aligns with U.S. kindergarten. Similarly, not all countries use "kindergarten" for the year prior to Grade 1, sometimes referring to it as "preschool." Due to these differences in terminology and school-entry age, studies describing multi-year preschool or kindergarten structures were standardized for "kindergarten" to refer to the year before Grade 1 and "preschool" to refer to the year before kindergarten.

Malleable Intervention Features, Instructional Strategies, and Mathematics Content. We extracted intervention information from each study, including the relationship of the intervention to the core instruction, instructional group size, dosage, and interventionist type. All studies were coded for various instructional strategies (see Figure 2) and early numeracy (see Figure 1); these are defined in the protocol (Nelson et al., 2025; https://osf.io/4m5gn/overview).

Screening Criteria Used to Identify Academic Risk. We identified how researchers screened each treatment group for academic risk and eligibility for the intervention study. To

Table 1. Description of Student Risk Groups Based on Screening Criteria.

Identified risk	Criteria for risk
MD High Risk	a. Students had a documented disability. b. Students were screened and performed at or below the 25th percentile on a mathematics measure.
MD Moderate Risk	a. Students were screened and performed above the 25th percentile in mathematics but below another specified threshold (e.g., 30th or 40th percentiles). b. Students were screened and identified for MD using alternative procedures such as teacher referral or a raw score on a test.
MD Very Low Risk	a. Students who were explicitly described as typically achieving with no risk factors, such as a low SES. b. Students were screened and performed above the 25th percentile.
SES Risk	a. Demographic data for the sample indicated that >75% of students were low SES. b. The study took place in a school or district with >75% of students being from low-income backgrounds.
No Screening Criteria	a. Studies did not report screening students for MD or did not report information about SES to determine risk.

Note. MD = mathematics difficulty; SES = socioeconomic status

create these screening criteria categories, we considered the screening practices that schools typically use to identify students as needing intervention, as well as established categories from other researchers (Lewis & Fisher, 2016) and previously published meta-analyses (Nelson & McMaster, 2019; Pereira et al., 2026). These categories and definitions are presented in Table 1 and include the following academic risk groups based on screening criteria: MD High Risk, MD Moderate Risk, MD Very Low Risk, SES Risk, and No Screening Criteria.

Briefly, MD High Risk referred to groups screened for mathematics-specific risk using performance below the 25th percentile or a documented disability; these criteria are commonly used by other researchers to identify high levels of risk for MD (e.g., Morgan et al., 2016). MD Moderate Risk referred to groups also screening for mathematics-specific risk, but with less stringent criteria compared with the MD High Risk group. SES Risk were groups identified as being at risk because of students' socioeconomic factors. Together, the MD High Risk, MD Moderate Risk, and SES Risk groups were all considered to have academic risk, with the MD groups having been screened specifically for learning difficulties in mathematics. Meanwhile, the MD Very Low Risk referred to groups that were screened for mathematics difficulties but found to be performing above a certain criterion and not identified as having MD. No Screening Criteria groups were identified as groups included in the intervention studies but with no reported screening criteria.

Data Analysis and Reliability

We calculated descriptive statistics (i.e., frequency, means, and standard deviations) to answer all research questions. For Research Questions 1 and 2, we calculated frequencies as raw data and percentages, as well as calculated ranges and averages. All analyses were conducted in Excel and R (version R-4.5.3). We used a variety of data visualization methods to report the results, including tables, bar graphs, and heat maps.

As part of the pre-registered meta-analysis, we conducted the double coding of all study variables for 25 studies (22%) to determine the reliability of the article coding procedures (see Benson, Farmer, et al., 2026, for additional details). For the current systematic review, we report reliability for only the variables used in the analyses. Interrater agreement was calculated as [agreements ÷ (agreements + disagreements) × 100] and ranged from 87.8% to 96.7%. Mean reliability across categories was as follows: basic study information (M = 94.4%), sample demographics (M = 87.8%), screening criteria (M = 96.7%), malleable intervention features (M = 91.6%), instructional strategies (M = 88.4%), and mathematics content (M = 88.6%).

Results

This systematic review includes 112 studies published between 2006 and 2025 from the pre-registered meta-analysis. The studies included journal articles (n=101; 90.2%), dissertations (n=8; 7.1%), conference proposals (n=2; 1.8%), and a book chapter (n=1; <1%). The studies included more than 21,000 child participants from 23 different countries (57 studies were conducted in the United States); gender representation was equivalent between males (43.3%) and females (43.5%), with the remaining participants' gender not defined. Participants in the United States were described as Hispanic (25.0%), White (24.2%), Black (20.1%), Multiracial (1.6%), Asian (1.5%), Other (1.7%), Native Hawaiian or Pacific Islander (<1%), and American Indian or Alaska Native (<1%) or not reported (25.4%). Among participants in the United States, 30.5% were identified as being from low-SES backgrounds, 14.8 % were identified as dual-language learners, and 5.6% were eligible for special education services. Studies conducted outside of the United States rarely reported information

about children's race/ethnicity or dual-language learner status. None of the 12 non-U.S. studies that reported information about SES used the same SES index; thus, summarizing SES for non-U.S. studies was impractical.

Across the 112 studies, there were 170 intervention groups that represented PreK only ($n=47$; 27.6%), kindergarten only ($n=65$; 38.2%), and Grade 1 only ($n=47$; 27.6%). There were few studies that grouped students across grade levels, including PreK and kindergarten ($n=8$; 4.7%) and kindergarten and Grade 1 ($n=3$; 1.8%).

RQ1: Screening Criteria Used to Identify Students for Mathematics Intervention Studies

First, we investigated the screening criteria and procedures that researchers used to identified PreK–1 students for inclusion in research studies. There was significant variability in the MD screening criteria and procedures that authors used, and therefore, the variability in the degree of risk that students were identified with was as follows: MD High Risk (14.1%), MD Moderate Risk (22.4%), MD Very Low Risk (5.9%), SES Risk (17.1%), and No Screening Criteria (40.6%). Thus, in total, 53.6% of treatment groups across the 112 early numeracy intervention studies were screened and identified as having academic risk, with 36.5% having been screened and identified as having mathematics-specific risk. Below, we discuss the results for each risk group.

Participant Risk Status as Determined by MD Screening Criteria. Across the 170 treatment groups, authors reported using screening criteria aligned with identifying a high likelihood of academic risk in mathematics (i.e., <25th percentile or disability) with 24 groups (14.1%). Most MD High Risk groups were identified using a percentile cutoff on a mathematics measure ($n=18$; 75.0%). Authors reported using performance below the 20th percentile for two groups and below the 25th percentile for 16 groups. Three additional groups were screened and identified as performing >1 *SD* or >1.5 SDs below the mean on a mathematics screening measure; we considered these cutoffs to be approximately the 16th and 7th percentiles, respectively. The remaining three groups were identified as MD High Risk because all students had a developmental delay ($n=1$) or a mathematics disability ($n=2$).

The most common screening criteria that authors used to identify academic risk aligned with MD Moderate Risk criteria ($n=38$ of 170 groups; 22.4%), which was the mathematics-specific screening criteria less stringent than MD High Risk. In a few instances, authors used a percentile on a screening test to identify academic risk ($n=5$; 13.2%), ranging from the 32nd to the 50th percentiles. Other screening criteria included using a raw score on a test that was not equated to a percentile cutoff ($n=12$; 31.6%), teacher referral only ($n=3$; 8.0%), and a combination of criteria such as teacher referral and a score on a test ($n=18$; 47.4%).

Ten intervention groups (5.9%) were identified as having MD Very Low Risk because the studies explicitly identified all students as having average or above-average performance in mathematics. Groups were identified as performing above the 25th ($n=6$), 30th ($n=1$), or 40th percentiles ($n=1$), with no requirement to perform *below* a specific percentile. Two groups were identified using a raw score, not percentile, on a screening measure.

In 24 studies, authors identified 29 treatment groups (17.1%) as having SES Risk, and therefore, being eligible for the intervention study. With these 24 studies, authors did not report any mathematics-specific screening criteria in addition to SES Risk. Authors reported enough information about the SES background of 14 of 29 groups to confirm that the study sample included a high proportion of students from low-SES backgrounds (i.e., >75% of the sample). All 14 studies were conducted in the United States, of which 11 were conducted in Head Start centers where authors reported federal poverty requirements for attendance. The remaining three studies reported free/reduced-price lunch (FRL) rates for the sample ranging from 87% to 100% eligible. Authors also reported sufficient information at the non-sample level for four groups to demonstrate that a high proportion of students was from low-SES backgrounds. These four studies were conducted in the United States, and authors reported school or district FRL rates ranging from 77% to 93%. Three groups were screened and identified by authors as having academic risk based on SES but had a lower proportion of students from low-SES backgrounds in the sample (44%–67%); these three studies were conducted outside of the United States, and authors reported contextual information that the low-SES rates of their sample were significantly higher than national averages. Eight groups did not include supporting information on how many students in the sample were from low-SES backgrounds.

Types of Measures Used to Identify MD. When authors screened students and identified groups as having mathematics-specific risk, the authors reported using a mathematics screening measure for 60 different groups across all 112 studies. The authors screened and identified 22 groups (36.7%) using norm-referenced measures (e.g., Test of Early Mathematics Ability and Woodcock-Johnson Mathematics Achievement Test). The remaining groups were screened and identified using well-established researcher-developed measures ($n=29$; 48.3%) or measures developed by the researchers for the purpose of their study ($n=9$; 15.0%). Established researcher measures included measures previously developed by authors that have often been used by author teams across many studies (e.g., the early numeracy measures developed and used by B. Clarke et al.;

the computation, concept, and application measures developed and used by L.S. Fuchs et al.).

RQ2: Patterns and Gaps by Screening Criteria

For RQ2, we identified patterns related to the level of evidence by grade level, malleable intervention features, instructional strategies, mathematics content, and screening criteria.

Grade Level: Patterns and Gaps. A summary of the results for screening criteria representation by grade level is presented in Supplemental Figure S2. The PreK and kindergarten intervention groups followed a similar pattern of risk. Most often, authors did not employ any screening procedures in PreK ($n=19$ groups; 42.2%) or kindergarten ($n=27$ groups; 40.3%). When authors reported screening procedures, the most common criteria to identify academic risk were SES Risk (PreK: $n=13$; 28.9%, K: $n=12$; 17.9%) and MD Moderate Risk criteria (PreK: $n=9$; 20%, K: $n=12$; 17.9%). A few groups at PreK ($n=3$; 6.7%) and kindergarten ($n=9$; 13.4%) were identified as having MD High Risk with performance below the 25th percentile on a screening measure. Grade 1 followed a slightly different pattern. The most common screening criteria that authors used included criteria to identify groups as MD Moderate Risk ($n=17$; 36.2%). This was followed closely by No Screening Criteria ($n=15$; 31.9%) and screening criteria aligned with MD High Risk ($n=12$; 25.5%). Only one Grade 1 study (2.1%) used SES Risk to identify students as being at academic risk. The mixed grade levels only identified students as having SES Risk ($n=3$; 27.3%) or did not use screening procedures ($n=8$; 72.7%).

Malleable Intervention Features: Patterns and Gaps. The intensity and duration of interventions varied by screening criteria. Children in the MD Moderate Risk group received the longest interventions on average (13.1 weeks and 37.2 sessions), followed by the MD High Risk group (9.8 weeks and 29.2 sessions). Children in the other risk groups received interventions that were shorter in duration, suggesting that on average, children screened and identified as having mathematics-specific risk received longer, more intensive interventions. See Supplemental Table S1 for the average duration, intensity, and total time spent in intervention across risk groups and instructional arrangement.

The length and intensity of interventions also varied by type of interventionist, with interventions conducted by a mix of research staff and teachers running for the longest duration (13.5 weeks and 46.5 sessions; see Supplemental Figure S3 for the results by type of interventionist). The cross-tabulation of screening criteria by interventionist type revealed that the highest total instructional time occurred in teacher-led interventions for groups identified with MD Moderate Risk criteria (1,505 min) and mixed-interventionist studies for MD Moderate Risk criteria (1,440 min; see Supplemental Figure S4). Across screening criteria, small-group instructional arrangements provided the most total instructional time for groups screened and identified as MD High Risk or MD Moderate Risk, whereas for SES Risk, the most prominent instructional arrangement was one-to-one. Interventionist type also corresponded with variation in dosage: Studies employing a mix of interventionists averaged the greatest duration (13.5 weeks and 46.5 sessions), followed by teacher-led interventions (11.8 weeks and 33.4 sessions), while research staff- and computer assisted interventions (CAI) were shorter on average (7.4 and 9.0 weeks, respectively). Supplemental Figure S4 provides a visualization of the instructional time across type of interventionist and screening criteria.

Instructional Strategies: Patterns and Gaps. A heatmap of instructional strategies disaggregated by screening criteria can be seen in Figure 2. In this figure, the proportion of intervention groups that implemented each instructional strategy for the indicated screening criteria is labeled in each cell. Note that many intervention groups incorporated more than one instructional strategy. The shading represents the proportion of strategy use, with darker shades indicating higher proportions (e.g., .76–1.00) and lighter shades indicating lower proportions (e.g., 0–.25).

Across screening criteria, guided or independent practice and frequent opportunities to respond (OTRs) were among the most commonly implemented strategies, indicating a strong emphasis on providing all students with repeated practice opportunities. Games were also widely used across all screening criteria, with particularly high representation among groups screened for and identified with mathematics-specific risk. Systematic instruction was another frequently implemented strategy, especially for students screened for and identified with mathematics-specific risk, aligning with recommended practices for supporting learners with MD (Fuchs et al., 2021). Visual representations and academic feedback were also commonly implemented across screening criteria. In contrast, cognitive strategy instruction and academic goal setting were less frequently used across screening criteria, indicating less emphasis on these strategies in the included studies.

Mathematics Content: Patterns and Gaps. A heatmap of mathematics content disaggregated by screening criteria can be seen in Figure 3. In this figure, the raw number of intervention groups that targeted the indicated content area for screening criteria is labeled in each cell. Note that many interventions targeted more than one skill area. The skills are clustered into three domains: number, relations, and operations.

Step 1	Step 2	Step 3
Identify the screening method	*Describe the screening criteria*	*Describe any exclusion criteria*
Documented disability	• Disability category • Individualized education program (IEP) goal in a specific area	• Disabilities categories or types (e.g., students with intellectual disabilities) • Performance below a certain percentile on an IQ-type test or other screening measure (e.g., reading) • Language requirements (e.g., English proficiency) • Students receiving other supports (e.g., Tier 3, specially designed instruction) • Students with behavioral support needs not able to be addressed during intervention • Students with scheduling conflicts • Students with chronic absenteeism
Percentile cutoff	• Report the percentile cutoff • Report the name of the test • Report if the test is norm-referenced or researcher-developed	
Raw score or percent correct	• Report the percent correct or raw score cutoff • Report the name of the test • Report if the test is norm-referenced or researcher-developed	
Teacher referral	• Report procedures teachers were instructed to follow • Report the evidence, criteria, or student characteristics considered • If students were identified using typical school procedures for intervention selection, report those procedures	
Socio-economic status	• Report if student, school, or district level data were used • Report the type of data used (e.g., free/reduced price lunch [FRL], household income, parents' highest level of education) • Report the SES level of the sample, if sample level data are not available report non-sample data (e.g., FRL district average)	
Other	• Identify the screening method, or combination of methods • Describe the screening and exclusion criteria with replicable details	
No screening method	• Report that no screening method was used to identify eligible students, and provide any other details (e.g., all students in the teachers' classrooms who opted into the study received intervention)	

Figure 3. Checklist for high-quality reporting of screening criteria in intervention research.
Note. This figure provides a checklist of the items to report for each screening method used to identify student participants for intervention studies.

In the domain of number skills, general counting ($n = 144$) and counting with one-to-one correspondence ($n = 106$) were commonly targeted across groups identified with different screening criteria. In general, counting (All Counting) was frequently covered with No Screening Criteria groups ($n = 56$). Skip counting was targeted infrequently across all groups (range = 0–5). The number skills domain had the highest frequency of the content targeted overall, with a representation of 565 times across intervention groups. In the relations domain, ordinality ($n = 14$) and numeral writing ($n = 31$) were targeted infrequently as compared with numeral identification ($n = 121$), number line skills ($n = 98$), and numeral or set comparison ($n = 90$). Notably, numeral identification was targeted with a high frequency in No Screening Criteria groups ($n = 45$).

Operations skills were targeted across studies with similar frequency to relations skills (312 and 354 times across intervention groups, respectively). Addition and subtraction skills were the most common operations skills targeted across all screening criteria ($n = 109$), with high representation in the MD Moderate Risk ($n = 34$) and No Screening Criteria ($n = 32$) groups. In contrast, equivalence ($n = 28$) and basic fact fluency ($n = 28$) were targeted much less frequently, particularly in groups screened and identified as MD Very Low Risk and SES Risk. Across all screening criteria and skills, skip counting ($n = 18$), ordinality ($n = 14$), basic fact fluency ($n = 28$), and equivalence ($n = 28$) were the least represented.

Discussion

This systematic review synthesized evidence from 112 PreK–1 early numeracy intervention studies included in a pre-registered meta-analysis (Author et al., 2026a) to examine how students were screened for academic risk and identify patterns and gaps in intervention features across screening criteria. Overall, the findings revealed substantial variability in how students were identified as being at academic risk in PreK–1 early numeracy intervention research. The findings also identified patterns and gaps in grade, malleable intervention features, instructional strategies, and mathematics content across screening criteria.

Patterns in Screening for Academic Risk and Research Directions

The results of this systematic review indicate that researchers are using a variety of screening criteria to identify PreK–1 students for early numeracy intervention studies. Some studies provided clear eligibility criteria, such as standardized cut scores, while others used broad labels, such as "low-achieving" or "struggling," without specifying how these groups were determined. Although many studies (40.6%)

did not report any screening criteria for their sample in terms of the students' need for mathematics intervention, authors reported clear screening criteria for many treatment groups (59.4%). This included authors using SES data (17.1%) and mathematics-specific criteria (42.4%); within these academic risk categories, authors reported a wide range of criteria. For example, authors reported low SES in U.S. studies, ranging from 77% to 100% of participants, while that in non-U.S. studies ranged from 44% to 67% of participants. With mathematics risk, the range of cut points to define MD spanned from approximately the 7th to the 50th percentiles. This result is aligned with MD literature, which indicates a general lack of consensus on MD identification criteria (Lewis & Fisher, 2016).

While it is not our intent to recommend a specific SES threshold or mathematics percentile cutoff that all researchers should strictly adhere to when conducting intervention research with students at risk, we provide the following key recommendations to improve future research. First, authors intending to make recommendations for students at risk for MD, when possible, should use either a norm-referenced or established researcher-developed measure to screen and identify eligible students for their intervention study. This aligns with real-world classroom practice in which an educator would identify students needing additional support, for example, based on fall screening data. Second, authors should report the test and percentile cutoff they used to identify eligible students. Although using raw scores and teacher referrals may be a more feasible approach to identifying students, using these methods makes it difficult for other researchers and practitioners to generalize the results of the research study to other populations. By enhancing reporting in journal articles, research can have a greater impact on classroom practice. Understanding what works and for whom depends on researchers clearly identifying their sample. Such findings align with field initiatives (e.g., Gersten et al., 2005) calling for greater transparency, specificity, and methodological rigor in MD identification.

Finally, if authors plan to use SES as an indicator of academic risk and eligibility for intervention, they should report sample-level SES information, including the source of information. In the current systematic review, when groups were identified as being at risk based on low SES ($n=29$), fewer than half (48.3%) reported sample-level demographic information about students' SES. This recommendation aligns with broader frameworks of generalization, including the classical generalization model and case-to-case transfer model, which emphasize the importance of understanding who was included in a study and how participants were selected (Cook & Cook, 2017). As Cook and Cook (2017) have noted, it is difficult for educators to apply research to their own students if information about samples and sampling is not reported clearly and consistently. Supplemental Table S1 includes a detailed checklist of items for authors to consider when reporting the screening criteria and specific details, as well as any exclusion criteria, for their intervention studies.

Patterns in Evidence by Screening Criteria and Research Directions

To inform the next steps for research and practice, we identified patterns and gaps by screening criteria and across grade level, malleable intervention features, instructional strategies, and mathematics content.

Grade Level. The results of the grade-level analysis revealed that fewer than half of the PreK (28.9%) and kindergarten (41.8%) groups were identified for the intervention study using specific mathematics screening criteria; however, in Grade 1, 66.0% of groups were identified using specific mathematics screening criteria. Generally, the representation of MD High Risk increased with each grade level from PreK (5.3%) to kindergarten (11.5%) to Grade 1 (25.5%), and the representation of No Screening Criteria decreased with each grade level (48.2%, 44.8%, and 31.9%, respectively), as did SES Risk (28.6%, 19.2%, and 2.1%, respectively). The increased use of mathematics-specific screening practices across grades is aligned with the results of previous research, indicating that the accuracy of identifying students as having MD improves with screening at the end of the kindergarten year or by screening multiple times in the kindergarten year (Morgan et al., 2009). We recommend that future researchers continue to use mathematics screening in PreK and kindergarten to identify students who may benefit most from intervention; with younger learners, researchers may consider using a higher percentile cutoff (e.g., 25th or 40th percentile). A higher cutoff allows researchers to intervene early with students who may be at a lower initial risk but may fall further behind over time, thus preventing persistent forms of MD with these students.

Malleable Intervention Characteristics. Regarding malleable intervention characteristics, the results indicated a pattern between screening criteria and the instructional group size. For example, while the one-to-one and small-group formats were common across groups, researchers only used whole-group instruction for children identified with SES Risk and No Screening Criteria. Mathematics researchers have increasingly recognized the role that intervention group size and within-group variability play in affecting student outcomes (Clarke et al., 2023; Sutherland, Kosty, et al., 2024). For example, Sutherland, Kosty, et al. (2024) found that intervention groups containing greater variability in students' mathematics skills at pretest made smaller gains from a Tier 2 kindergarten mathematics intervention, and this pattern was amplified in larger intervention groups. Future systematic review and meta-analytic research should

investigate the role of intervention group size and within-group variability on mathematics outcomes to inform intervention planning and resource allocation within schools.

Interestingly, children with SES Risk received markedly more total instructional time for CAIs as compared with any other screening criterion. One possible explanation for this pattern is that low-SES schools often have higher rates of novice teachers and teacher turnover (Carver-Thomas & Darling-Hammond, 2019), which may make sustainable intervention options such as CAI more feasible. Another key finding related to intervention features was that while 71 groups in total received interventions led by researchers, only 38 received interventions from teachers. This finding is particularly timely with the recent attention given to considerations for the long-term feasibility and relevance of research conducted in schools (Northern & Opp, 2026). Future researchers may consider partnering with schools prior to establishing an intervention research study to create a plan for training teachers to implement the intervention as part of the study or transitioning implementation to teachers following the research study to ensure the long-term sustainability of intervention practices.

Mathematics Content: Patterns and Gaps. Early numeracy interventions for PreK–1 students primarily focus on foundational number skills, with strong coverage in the No Screening Criteria group. Interventions for students screened for and identified with MD Moderate Risk also frequently targeted core number and operations skills, particularly addition and subtraction, suggesting that research focuses on foundational skill-building for children with identified MD. In contrast, critical content areas that support later mathematics learning, such as skip counting ($n = 18$), basic fact fluency ($n = 28$), and equivalence ($n = 28$), were rarely targeted. These are key foundational skills that may benefit from earlier intervention. Equivalence, in particular, is essential for promoting algebraic thinking and the conceptual understanding of equations and operations.

The distribution of content across screening criteria reflects the broader patterns in the sample: The MD Moderate Risk and No Screening Criteria populations are most highly represented, with less research for the MD High Risk and SES Risk populations. Future research that expands interventions to include less frequently addressed but developmentally essential skills, especially for populations with fewer research studies, could strengthen our understanding of which early intervention practices promote durable and sustained outcomes in mathematics.

Instructional Strategies: Patterns and Gaps. The results of the instructional strategy analysis indicated that when certain strategies were investigated in PreK–1 early numeracy intervention, they were similarly investigated across screening criteria. For example, the heat map shows a pattern of investigating the effectiveness of games, academic feedback, representations, frequent opportunities to respond, guided and independent practice, and explicit and/or systematic instruction across most screening criteria. The emphasis on these instructional practices is similar to the results of a review of 75 meta-analyses focused on high-leverage practices (HLPs) for students with or at risk of disabilities across academics and behavior (Nelson et al., 2022). The authors of the HLP review found that commonly investigated HLPs included promoting engagement (e.g., frequent responses and active participation), using explicit instruction (e.g., providing practice opportunities), systematic instruction, and using scaffolding (e.g., visual support; Nelson et al., 2022). Although generally well represented, concrete and visual representations were emphasized more often for students in the MD High Risk group. Given the strong evidence supporting the effectiveness of multiple representations in early numeracy instruction (Fuchs et al., 2021), future intervention studies can further incorporate mathematical representations to support learning for all at-risk students.

In contrast, cognitive strategy instruction, problem solving, and academic goal setting appeared less frequently in the heatmap. This pattern was unexpected, given that cognitive factors have been identified as significant predictors of mathematics outcomes (Geary, 2011; Mazzocco, 2005; Swanson et al., 2021) and self-monitoring is an important component of mathematics learning (Sutherland, Lussier, et al., 2024). Future intervention research studies may continue to explore the effectiveness of these strategies for different groups of students.

Implications for Practice

The findings from this review have several implications for improving early numeracy instruction and interventions for students at academic risk. Across studies, there was substantial variability in how students were identified for intervention, with over 40% of intervention groups lacking a clearly defined criterion for screening children for being at academic risk. Instructional strategies that have been tested and shown to be effective with clearly defined populations, such as students with MD, may be used more frequently by interventionists seeking approaches for these populations. On the other end of the spectrum, approaches tested with class-wide or broadly defined student populations such as SES-based risk may be most appropriately selected for enhancing Tier 1 instruction or preventative approaches until more evidence is available regarding their use with at-risk populations. The findings also highlight the importance of aligning intervention intensity and implementation with student needs. Practitioners should consider systematic decision-making around intervention intensity based on contextual factors in their setting and closely monitor the

progress for students at risk. In particular, students at risk due to SES who may attend under-resourced schools should be monitored closely for their progress, maintenance, and generalization, especially when interventions are primarily delivered via CAI.

With regard to mathematics content, early intervention research has largely focused on foundational number skills, which is in alignment with grade-level standards and developmental expectations. However, several content areas critical for the development of multiplicative reasoning, algebraic reasoning, and overall mathematics fluency (i.e., skip counting, ordinality, fact fluency, and equivalence) were infrequently targeted. It is important to introduce and practice these concepts to support success in mathematics in later grades.

Well-established, evidence-based teaching practices were strongly represented in this review, including guided practice, frequent opportunities to respond, use of visual representations, and instructional games. Practitioners should prioritize strategies with a robust evidence base while recognizing that less-researched practices such as cognitive strategy instruction and goal setting may support independence and generalization in mathematical thinking. Finally, the malleable intervention features in this corpus of studies raise important considerations for the feasibility and implementation of interventions in authentic school contexts. Because many interventions were delivered by researchers, practitioners must carefully consider the contextual variables and feasibility of approaches when implementing interventions.

Limitations

This systematic review has the following limitations. First, because we aligned this study with a pre-registered meta-analysis, we only included studies in this systematic review that met specific design criteria. The results of this systematic review may not be generalizable to the broader PreK–1 early numeracy intervention literature. Second, our review focuses on PreK–1 students. Because many students may not yet be identified as being at risk, the patterns that we observed in this systematic review may differ when studies in later grades are included. Future researchers can examine patterns in evidence according to screening criteria across other grade levels. Finally, although the screening criteria used in this systematic review are based on previous research, the groups established from these criteria are not mutually exclusive, limiting our ability to identify more precise patterns based on the degree of students' risk. For example, the MD Moderate Risk group—when students were identified with only teacher referral, for example—may have included students who met the criteria for MD High Risk (i.e., performance <25th percentile) had they been screened. This limitation reflects real-world differences in screening approaches and cannot be addressed within a systematic review unless all studies report using the same type of screening criteria. Although the screening criteria categories are not mutually exclusive, researchers can still learn from these findings and improve on their own scholarship. For example, researchers who are interested in students from low-SES backgrounds can look at the results of this systematic review and determine the areas within PreK–1 in which early numeracy interventions may be underexplored and their work can be expanded.

Conclusion

This systematic review highlights the variability in how PreK–1 students are identified for early numeracy intervention, limiting the interpretability and application of findings across contexts. At the same time, patterns in intervention features, content, and instructional strategies suggest both areas of strength and critical gaps in the evidence base. Advancing the field will require greater transparency, more consistent identification practices, and expanded research across under-represented risk groups to support more equitable and impactful mathematics interventions.

ORCID iDs

Gena Nelson https://orcid.org/0000-0002-7369-0132
Soyoung Park https://orcid.org/0000-0001-6178-6818
Sarah Quinn https://orcid.org/0000-0002-7309-0110
Taylor Lesner https://orcid.org/0000-0003-2797-1491

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the U.S. Department of Education, Institute of Education Sciences grant R324A240004 to the University of Oregon. The opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the Institute of Education Sciences.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Drs. Clarke, Lesner, Shanley, and Sutherland are authors of studies included in this systematic review. As such, steps were taken to ensure reliability of the screening and article coding stages.

Supplemental Material

Supplementary material for this article is available on the Learning Disability Quarterly website at http://journals.sagepub.com/home/ldq.10

References

A list of the 112 studies in this systematic review can be found in the online supplementary files.

Benson, N. F., Farmer, R. L., Maki, K. E., & Hajovsky, D. B. (2026). Redefining specific learning disabilities: An attempt

to bridge research, policy, and practice through expert consensus. Journal of School Psychology, 116, 101549. https://doi.org/10.1016/j.jsp.2026.101549

Bryant, D. P., Bryant, B. R., Gersten, R. M., Scammacca, N. N., Funk, C., Winter, A., Shih, M., & Pool, C. (2008). The effects of tier 2 intervention on the mathematics performance of first-grade students who are at risk for mathematics difficulties. *Learning Disability Quarterly*, *31*(2), 47–63. https://doi.org/10.2307/20528817

Burchinal, M., McCartney, K., Steinberg, L., Crosnoe, R., Friedman, S. L., McLoyd, V., Pianta, R., & NICHD Early Child Care Research Network. (2011). Examining the Black–White achievement gap among low-income children using the NICHD study of early child care and youth development. *Child Development*, *82*(5), 1404–1420. https://doi.org/10.1111/j.1467-8624.2011.01620.x

Carver-Thomas, D., & Darling-Hammond, L. (2019). The trouble with teacher turnover: How teacher attrition affects students and schools. *Education Policy Analysis Archives*, *27*, 36. https://doi.org/10.14507/epaa.27.3699

Cary, M. S., Kennedy, P. C., Shanley, L., & Clarke, B. (2021). Learning gains from the KinderTEK® iPad math program: Does timing of a preventative intervention matter?. *Journal of Special Education Technology*, *36*(4), 321–335. https://doi.org/10.1177/0162643420928336

Charitaki, G., Tzivinikou, S., Stefanou, G., & Soulis, S. G. (2021). A meta-analytic synthesis of early numeracy interventions for low-performing young children. *SN Social Sciences*, *1*(5), 105. https://doi.org/10.1007/s43545-021-00094-w

Clarke, B., Cil, G., Smolkowski, K., Sutherland, M., Turtura, J., Doabler, C. T., Fien, H., & Baker, S. K. (2020). Conducting a cost-effectiveness analysis of an early numeracy intervention. *School Psychology Review*, *49*(4), 359–373. https://doi.org/10.1080/2372966X.2020.1761236

Clarke, B., Doabler, C. T., Sutherland, M., Kosty, D., Turtura, J., & Smolkowski, K. (2023). Examining the impact of a first grade whole number intervention by group size. *Journal of Research on Educational Effectiveness*, *16*(2), 326–349. https://doi.org/10.1080/19345747.2022.2093299

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/ldrp.12132

Davis-Kean, P. E., Domina, T., Kuhfeld, M., Ellis, A., & Gershoff, E. T. (2022). It matters how you start: Early numeracy mastery predicts high school math course-taking and college attendance. *Infant and Child Development*, *31*(2), e2281. https://doi.org/10.1002/icd.2281

Dennis, M. S. (2015). Effects of tier 2 and tier 3 mathematics interventions for second graders with mathematics difficulties. *Learning Disabilities Research & Practice*, *30*(1), 29–42. https://doi.org/10.1111/ldrp.12051

Dietrichson, J., Filges, T., Seerup, J. K., Klokker, R. H., Viinholt, B. C., Bøg, M., & Eiberg, M. (2021). Targeted school-based interventions for improving reading and mathematics for students with or at risk of academic difficulties in Grades K-6: A systematic review. *Campbell Systematic Reviews*, *17*(2), e1152. https://doi.org/10.1002/cl2.1152

Doabler, C. T., Clarke, B., Kosty, D., Maddox, S. A., Smolkowski, K., Fien, H., Baker, S., & Kimmel, G. L. (2021). Kindergarteners at risk for severe mathematics difficulties: Investigating tipping points of core mathematics instruction. *Journal of Learning Disabilities*, *54*(2), 97–110. https://doi.org/10.1177/0022219420972185

Filiz, T., & Güneş, G. (2023). A systematic review of instructional interventions applied to primary school students with mathematics learning difficulties. *International Journal of Education and Literacy Studies*, *11*(4), 281–296. https://doi.org/10.7575/aiac.ijels.v.11n.4p.281

Fuchs, L. S., Newman-Gonchar, R., Schumacher, R., Dougherty, B., Bucka, N., Karp, K. S., Woodward, J., Clarke, B., Jordan, N. C., Gersten, R., Jayanthi, M., Keating, B., & Morgan, S. (2021). *Assisting students struggling with mathematics: Intervention in the elementary grades*. National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. https://ies.ed.gov/ncee/wwc/PracticeGuide/26

Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *Journal of Developmental & Behavioral Pediatrics*, *32*(3), 250–263. https://doi.org/10.1097/DBP.0b013e318209edef

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, *71*(2), 149–164. https://doi.org/10.1177/001440290507100202

Jitendra, A. K., Alghamdi, A., Edmunds, R., McKevett, N. M., Mouanoutoua, J., & Roesslein, R. (2021). The effects of tier 2 mathematics interventions for students with mathematics difficulties: A meta-analysis. *Exceptional Children*, *87*(3), 307–325. https://doi.org/10.1177/0014402920969187

Lewis, K. E., & Fisher, M. B. (2016). Taking stock of 40 years of research on mathematical learning disability: Methodological issues and future directions. *Journal for Research in Mathematics Education*, *47*(4), 338–371. https://doi.org/10.5951/jresematheduc.47.4.0338

Lysenko, L. V., Rosenfield, S., Dedic, H., Savard, A., Idan, E., Abrami, P. C., Wade, A., & Naffi, N. (2016). Using interactive software to teach foundational mathematical skills. *Journal of Information Technology Education: Innovations in Practice*, *15*, 19–34. http://www.jite.org/documents/Vol15/JITEv15IIPp019-034Lysenko2154.pdf. https://doi.org/10.28945/3395

MacDonald, A., & Murphy, S. (2021). Mathematics education for children under four years of age: A systematic review of the literature. *Early Years*, *41*(5), 522–539. https://doi.org/10.1080/09575146.2019.1624507

Mazzocco, M. M. M. (2005). Challenges in identifying target skills for math disability screening and intervention. *Journal of Learning Disabilities*, *38*(4), 318–323. https://doi.org/10.1177/00222194050380040701

Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school age years. *Annals of Dyslexia*, *53*, 218–253. https://doi.org/10.1007/s11881-003-0011-7

Monei, T., & Pedro, A. (2017). A systematic review of interventions for children presenting with dyscalculia in primary schools. *Educational Psychology in Practice*, *33*(3), 277–293. https://doi.org/10.1080/02667363.2017.1289076

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2016). Who is at risk for persistent mathematics difficulties in the United States?. *Journal of learning disabilities*, *49*(3), 305–319. https://doi.org/10.1177/0022219414553849

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning disabilities*, *42*(4), 306–321. https://doi.org/10.1177/0022219408331037

National Center for Education Statistics & US Department of Education. (2025). *Assessments—Mathematics*. https://nces.ed.gov/nationsreportcard/mathematics

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence*. https://www.nctm.org/store/Products/Curriculum-Focal-Points-for-Prekindergarten-through-Grade-8-Mathematics-A-Quest-for-Coherence/

National Research Council. (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. National Academies Press. https://ui.adsabs.harvard.edu/abs/2009nap..book12519N/abstract

Nelson, G., Boedeker, P., Lesner, T., & Clarke, B. (2025). Early math meta-analysis (preK-1). *OSF*. https://doi.org/10.17605/OSF.IO/4M5GN

Nelson, G., Boedeker, P., Lesner, T., & Clarke, B. (2026). Early math meta-analysis (preK-1). Preregistration. *OSF Registries*. https://osf.io/4m5gn/registrations

Nelson, G., Boedeker, P., Lesner, T., & Clarke, B., Sutherland, M., Shanley, L., Park, S., Kim, G., Quinn, S., Ehrman, P., Brafford, T., Fainstein, D., & Drake, K. R. (2026). What works in early mathematics intervention? Findings from a meta-analysis of 112 preschool through grade 1 studies. *Unpublished manuscript, Center on Teaching and L*

Nelson, G., Cook, S. C., Zarate, K., Powell, S. R., Maggin, D. M., Drake, K. R., Kiss, A. J., Ford, J. W., Sun, L., & Espinas, D. R. (2022). A systematic review of meta-analyses in special education: Exploring the evidence base for high-leverage practices. *Remedial and Special Education*, *43*(5), 344–358. https://doi.org/10.1177/07419325211063491

Nelson, G., & McMaster, K. L. (2019). The effects of early numeracy interventions for students in preschool and early elementary: A meta-analysis. *Journal of Educational Psychology*, *111*(6), 1001–1022. https://doi.org/10.1037/edu0000334

Nelson, G., & Powell, S. R. (2018). A systematic review of longitudinal studies of mathematics difficulty. *Journal of Learning Disabilities*, *51*(6), 523–539. https://doi.org/10.1177/0022219417714773

Northern, A., & Opp, A. (2026). *Reimagining the Institute of Education Sciences: A strategy for relevance and renewal*. U.S. Department of Education. https://ies.ed.gov/ies/2026/02/reimagining-ies

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . .Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906. https://doi.org/10.1016/j.ijsu.2021.105906

Pereira, A. R., Van Belle, L., Wijns, N., Van Den Noorgate, W., & De Smedt, B. (2026). *It counts. A. multilevel meta-analysis on early numeracy interventions*. Unpublished manuscript, Psychology and Educational Sciences, KU Leuven.

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Schnepel, S., & Aunio, P. (2022). A systematic review of mathematics interventions for primary school students with intellectual disabilities. *European Journal of Special Needs Education*, *37*(4), 663–678. https://doi.org/10.1080/08856257.2021.1943268

Sutherland, M., Kosty, D., Lesner, T., Hermida, J., Smolkowski, K., Doabler, C. T., & Clarke, B. (2024). Does variability within tier 2 mathematics intervention groups affect students' response to intervention? *Exceptional Children*, *90*(3), 331–346. https://doi.org/10.1177/00144029231220307

Sutherland, M., Lussier, C., Nelson, G., Suhr, M. P., Fong, J., Turtura, J., & Clarke, B. (2024). A quantitative systematic literature review of self-monitoring components within mathematics instruction and intervention. *Exceptional Children*, *90*(3), 235–254. https://doi.org/10.1177/00144029241230322

Svane, R. P., Willemsen, M. M., Bleses, D., Krøjgaard, P., Verner, M., & Nielsen, H. S. (2023). A systematic literature review of math interventions across educational settings from early childhood education to high school. *Frontiers in Education*, *8*, Article 1229849. https://doi.org/10.3389/feduc.2023.1229849

Swanson, H. L., Arizmendi, G. D., & Li, J.-T. (2021). Working memory growth predicts mathematical problem-solving growth among emergent bilingual children. *Journal of Experimental Child Psychology*, *201*, Article 104988. https://doi.org/10.1016/j.jecp.2020.104988

Tan, C. Y. (2024). Socioeconomic status and student learning: Insights from an umbrella review. *Educational Psychology Review*, *36*(4), Article 100. https://doi.org/10.1007/s10648-024-09929-3

Toll, S. W., & Van Luit, J. E. (2012). Early numeracy intervention for low-performing kindergartners. *Journal of Early Intervention*, *34*(4), 243–264. https://doi.org/10.1177/1053815113477205

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, *36*, 541–561. https://doi.org/10.1080/02796015.2007.12087917

Williams, R., Citkowicz, M., Miller, D. I., Lindsay, J., & Walters, K. (2022). Heterogeneity in mathematics intervention effects: Evidence from a meta-analysis of 191 randomized experiments. *Journal of Research on Educational Effectiveness*, *15*(3), 584–634. https://doi.org/10.1080/19345747.2021.2009072

Yazıcıoğlu, T., & Akdal, D. (2023). An analysis of the effects of the Early Mathematics Intervention Program on early mathematics skills of pre-school children at risk. *Participatory Educational Research*, *10*(1), 237–250. https://doi.org/10.17275/per.23.13.10.1