AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

ISSN: 0022-0663

2024, Vol. 116, No. 1, 1–19
https://doi.org/10.1037/edu0000820

Unique and Combined Effects of Quantitative Mathematical Language and Numeracy Instruction Within a Picture Book Intervention: A Registered Report

David J. Purpura[1], Connor D. O'Rear[2], Alexa Ellis[3], Jessica A. R. Logan[4], Lauren Westerberg[1], Patrick Ehrman[1], Yemimah A. King[5], Mackenna Vander Tuin[6], Inga Nordgren[1], Kirsten Anderson[1], Jimena Cosso[7], Erica Zippert[8], Amy R. Napoli[9], Caroline Byrd Hornburg[10], Sara A. Schmitt[11], and Jennifer Dobbs-Oates[1]

[1] Department of Human Development and Family Science, Purdue University
[2] Institute of Research, Innovation, and Strategy, University of Notre Dame
[3] Department of Human Development and Family Studies, University of Alabama
[4] Department of Special Education, Vanderbilt University
[5] Department of Early Childhood and Elementary Education, Georgia State University
[6] Department of Special Education, University of Texas at Austin
[7] Department of Educational Psychology, Counseling, and Special Education, Pennsylvania State University
[8] Learning Research and Development Center, University of Pittsburgh
[9] Department of Child, Youth, and Family Studies, University of Nebraska
[10] Department of Human Development and Family Science, Virginia Tech
[11] Department of Special Education and Clinical Science, University of Oregon

Children's early understanding of mathematics provides a foundation for later success in school. Identifying ways to enhance mathematical instruction is crucial to understanding the ideal ways to promote academic success. Previous work has identified mathematical language (i.e., the words and concepts related to early mathematical development such as more, same, or similar) as a key mechanism that can be targeted to improve children's development of early numeracy skills (e.g., counting, cardinality, and addition). Current recommendations suggest a combination of numeracy instruction and quantitative language instruction to promote numeracy skills. However, there is limited direct support of this recommendation. The goal of the proposed study is to compare the unique and combined effects of each type of instruction on children's numeracy skills in the context of picture book reading. We randomly assigned 234 children (ages 3–5) to one of four conditions where they worked with trained project staff who read picture books targeting: (a) quantitative language only (e.g., more or less), (b) numeracy only (e.g., cardinality, addition), (c) combined [quantitative

This article was published Online First November 2, 2023.

Alexa Ellis https://orcid.org/0000-0001-7481-5788
Lauren Westerberg https://orcid.org/0000-0003-0663-5501
Jimena Cosso https://orcid.org/0000-0001-8313-8547
Erica Zippert https://orcid.org/0000-0002-1439-6574
Caroline Byrd Hornburg https://orcid.org/0000-0001-9563-859X
Jennifer Dobbs-Oates https://orcid.org/0000-0002-1215-2672

The research reported here was supported by the National Science Foundation (Award 1749294) to Purdue University. The opinions expressed are those of the authors and do not represent the views of the National Science Foundation. Some of the books designed for this study are available for purchase on Amazon. Although no profits/royalties are directly obtained by any of the authors, any profits from the sale of the books go toward a university-held account for David J. Purpura to use to put toward the development of new books or the purchase of copies of the books to give away.

David J. Purpura served as lead for conceptualization, funding acquisition, investigation, resources, supervision, and writing–review and editing and served in a supporting role for data curation, formal analysis, methodology, and project administration. Connor D. O'Rear served in a supporting role for data curation, investigation, project administration, and writing–review and editing. Alexa Ellis served in a supporting role for formal analysis, investigation, methodology, project administration, and writing–review and editing. Jessica A. R. Logan served as lead for formal analysis and served in a supporting role for data curation, methodology, validation, writing–original draft, and writing–review and editing. Lauren Westerberg served as lead for data curation, project administration, and supervision and served in a supporting role for conceptualization, investigation, writing–original draft, and writing–review and editing. Patrick Ehrman served in a supporting role for data curation, investigation, project administration, supervision, and writing–review and editing. Yemimah A. King served in a supporting role for conceptualization, investigation, resources, writing–original draft, and writing–review and editing. Mackenna Vander Tuin served in a supporting role for data curation, investigation, project administration, and writing–review and editing. Inga Nordgren served in a supporting role for investigation, supervision, and writing–review and editing. Kirsten Anderson served in a supporting role for investigation and writing–review and editing. Jimena Cosso served in a supporting role for investigation, writing–original draft, and writing–review and editing. Erica Zippert served in a supporting role for writing–original draft and writing–review and editing. Amy R. Napoli served in a supporting role for conceptualization, writing–original draft, and writing–review and editing. Caroline Byrd Hornburg served in a supporting role for conceptualization, writing–original draft, and writing–review and editing. Sara A. Schmitt served in a supporting role for conceptualization, writing–original draft, and writing–review and editing. Jennifer Dobbs-Oates served in a supporting role for conceptualization, writing–original draft, and writing–review and editing. David J. Purpura, Connor D. O'Rear, and Alexa Ellis contributed equally to writing–original draft.

Correspondence concerning this article should be addressed to David J. Purpura, Department of Human Development and Family Science, Purdue University, 1200 West State Street, Room 223, West Lafayette, IN 47907, United States. Email: purpura@purdue.edu

language + numeracy], or (d) nonnumerical (active control) picture books. Results revealed no significant effects of the quantitative language only or numeracy only conditions, but mixed effects of the combined condition. These findings indicate that more work is needed on how mathematical language and numeracy instruction should best be delivered to preschool children.

Educational Impact and Implications Statement
Significant research has highlighted the importance of quantitative language in children's early mathematics development. The findings from this study raise questions as to how to best implement such interventions. Namely, in contrast to prior work, the current interventions targeting one of the two domains did not exhibit positive results. However, mixed effects were found for the condition that combined both domains. Future work on how to best structure combined instruction is needed.

Keywords: quantitative mathematical language, numeracy, picture book intervention, preschool, mathematics

Children's early mathematical skills provide the foundation on which later mathematical skills are built (Aunola et al., 2004; Purpura et al., 2013). However, other academic and cognitive domains are also important for this development. Language, in particular, is important for children's understanding of mathematics (LeFevre et al., 2010; Purpura et al., 2011). Recent evidence indicates that understanding of mathematical language (e.g., more, few, less, before, after) is a strong predictor of early mathematics development, even more so than broad language skills (Purpura & Logan, 2015; Toll & Van Luit, 2014). Children's mathematical language knowledge at the start of preschool is a better indicator of later risk for mathematics difficulties than initial mathematics skills (Purpura, Day, et al., 2017). Intervention work has also demonstrated that directly instructing mathematical language promotes the development of early numeracy skills (e.g., counting, cardinality, and addition; Purpura, Napoli, et al., 2017).

Though the concept of introducing mathematical language to children may appear relatively simple, merely encouraging parents and teachers to use more mathematical language may not be sufficient for improving children's skills because of the breadth of mathematical language concepts (Boonen et al., 2011; Lansdell, 1999). Notably, providing children with diverse mathematical talk (i.e., using a greater variety of mathematical words), particularly if that talk is above children's current developmental abilities (e.g., using word problems to discuss calculation when children are still learning to count), may actually cause confusion and can be negatively related to their understanding of mathematics (e.g., Boonen et al., 2011). Thus, introducing children to more narrowly targeted mathematical language that aligns with specific mathematics skills may be most effective. Moreover, research and policy recommendations suggest combining mathematical language with mathematics instruction (Chard et al., 2008; Clements & Sarama, 2011). However, it remains unclear if there is a benefit to joint instruction combining exact numeracy instruction with quantitative mathematical language rather than instruction in either domain alone. The purpose of this study is to examine the unique and combined effects of quantitative mathematical language and exact numeracy instruction for promoting preschool children's quantitative mathematical language and numeracy skills within the context of a picture book reading intervention.

Language and Mathematics Development

There is a strong relation between mathematics and language skills that appears early in children's academic development (Hooper et al., 2010; Purpura et al., 2011; Romano et al., 2010). Children use language skills to develop a more complex understanding of mathematics (Spelke, 2003). Relatedly, difficulties with mathematics, literacy, and language often co-occur (Koepke & Miller, 2013; C. Lewis et al., 1994) and having difficulties in both mathematics and language is associated with lower mathematics performance than having difficulties in mathematics alone (Hanich et al., 2001; Jordan & Hanich, 2000).

Environmental contexts also contribute to the relation between language and mathematics (e.g., Jordan & Levine, 2009; Jordan et al., 1994; Starkey et al., 2004). Socioeconomic status (SES)-related differences in language skills are thought to be one reason for the SES-related differences observed in mathematics achievement (Jordan & Levine, 2009; Jordan et al., 1994; Starkey et al., 2004). Despite this explanation, research has suggested that improving children's general language skills does not necessarily benefit their mathematics knowledge (Jordan et al., 2012). A likely explanation for this lack of improvement is that the language targeted through general language interventions often includes nonmathematical words, or words that are used differently in nonmathematical contexts than those used in mathematical instruction (Harmon et al., 2005). Thus, language instruction specifically focused on mathematical language may be one way to use the environmental context to improve children's mathematical knowledge.

Mathematical Language

Distinct from mathematical knowledge (specifically, exact numeracy: counting, cardinality, addition), children's mathematical language refers to their understanding of key terms and concepts used in early mathematics such as more, few, similar, before, and after (Purpura et al., 2019). These terms and concepts are often couched within a framework of relational language (Barner & Bachrach, 2010; Loewenstein & Gentner, 2005) where the links between quantities or shapes and space are approximate and not exact. Two different types of this relational, or mathematical, language have been identified as relevant to children's early mathematical learning: quantitative language (Barner et al., 2009; Purpura et al., 2019) and spatial language (Pruden et al., 2011; Ramani et al., 2014).

Evidence suggests that quantitative mathematical language, specifically, drives the causal relation between mathematical language and numeracy growth (Purpura et al., 2019, 2021). Thus, the link between quantitative mathematical language (not spatial language) and numeracy is the focus of the current study.

Quantitative language includes terms that denote relative quantity such as more than, less than, many, and fewer (some of the terms have also been labeled quantifiers; Barner, Chow, & Yang, 2009). Understanding these terms allows children to make and describe comparisons between groups or numbers, such as knowing that the term more can mean an increase in quantity ("give me more") or can be used in comparative statements ("she has more than I do"). Using these comparative terms may help children refine their initial, approximate understanding of numbers (Barner, Chow, & Yang, 2009). Being able to compare and identify the relations between quantities using quantitative mathematical language may help draw children's attention to set size and refine their understanding of exact number.

Previous Numeracy and Mathematical Language Interventions

A strong emphasis on mathematical language is believed to be a critical component of the success of early mathematics instruction (Chard et al., 2008; Clements & Sarama, 2011; National Council of Teachers of Mathematics, 2006); however, to date, only a few intervention studies with 3–5 years old have focused on investigating the direct connection between children's quantitative mathematical language understanding and their numeracy skills (Purpura et al., 2021; Purpura, Napoli, et al., 2017), though others have separately investigated the effects of combined (mathematical language + numeracy) instruction (e.g., Green et al., 2018; Jennings et al., 1992; Young-Loveridge, 2004) or exact numeracy instruction alone (i.e., counting and cardinality interventions; e.g., Gibson et al., 2020; Mix et al., 2012; O'Rear & McNeil, 2019). However, given the divergent designs of these prior studies, it is not possible to directly contrast the unique and combined impacts of quantitative mathematical language and numeracy instruction. There is a critical need to understand how integration of quantitative mathematical language instruction and numeracy instruction can be used to support children's early numeracy success and to inform future development of effective instructional methods. In the current study, we conducted an intervention to examine the unique and combined effects of both types of instruction within the context of picture book reading, as picture books are a central part of the early learning environment that is relevant to both the home and school settings.

Mathematical Language and Numeracy Instruction Alone

A broad range of work has demonstrated the effectiveness of improving elementary school-aged children's numeracy skills through direct numeracy instruction using curricula and intervention materials (Hassinger-Das et al., 2015; Powell & Driver, 2015), and numeracy skills for 3–5 years old through more targeted efforts including using picture books (Gibson et al., 2020; Green et al., 2018; Jennings et al., 1992; Mix et al., 2012; O'Rear & McNeil, 2019; Purpura et al., 2021; Purpura, Napoli, et al., 2017; Young-Loveridge, 2004) mining mathematical language instruction offers some insight into how mathematical language is related to numeracy skills. Purpura, Napoli, et al. (2017) assigned 3- to 5-year-old children from Head Start centers to either a mathematical language intervention or a business-as-usual control condition. Children in the mathematical language group participated in an 8-week dialogic reading program in small groups where interventionists elaborated on mathematical content within the book (e.g., "How do we know there were a lot and not a few?") or used prompts to incorporate quantitative and spatial mathematical language (e.g., "There are just a few people in the stands." or "Is the roller coaster near the ground?"). To focus on the role of mathematical language specifically, the language used in this study did not include a focus on exact numeracy or counting, and interventionists were explicitly trained to not use numeracy concepts during the instructional periods. Children in the mathematical language picture book condition outperformed children in the control condition on a measure of mathematical language and on a broad measure of early numeracy skills. Broadly targeting mathematical language instruction (including quantitative and spatial mathematical terms) is one way to promote children's early numeracy skills.

Follow-up analyses to the above work (Purpura et al., 2019) suggest the type of mathematical language instruction may matter. In that study, the authors found that both the experimental condition and the control condition showed similar improvements in spatial mathematical language, but only the mathematical picture book condition demonstrated improvement in quantitative mathematical language. As the mathematical picture book condition also saw a greater improvement in numeracy skills, the authors reasoned that the advantages in numeracy scores for the experimental condition are most likely due to the increased quantitative mathematical language skills for this condition.

In a more targeted intervention study (Purpura et al., 2021), the specific impact of quantitative mathematical language on children's performance was evaluated. Specifically, parents of 3- to 5-year-old children were randomly assigned to read picture books that included a focus on quantitative mathematical language or nonmathematical picture books. Parents were asked to read three times a week for 4 weeks, and each book included dialogic reading prompts to increase engagement. In the experimental condition, some of the prompts focused on quantitative mathematical language (e.g., "Who has more pillows: Bear or Benjamin?"). At an immediate posttest, children in the experimental condition outperformed children in the control condition on a measure of their quantitative mathematical language as well as on a measure of numeracy. At an 8-week delayed posttest, children in the experimental condition continued to outperform children in the control condition. Thus, focusing on quantitative language is an effective way to improve children's early numeracy skills.

Combining Mathematical Language and Numeracy Instruction

Some numeracy interventions have also demonstrated positive effects when exact numeracy instruction incorporates mathematical language. Jennings et al. (1992) incorporated picture books during mathematical instruction with the goal of increasing kindergarten (5–6 years old) children's mathematics achievement. Kindergarten teachers read mathematical picture books with children before expanding upon the story using mathematically focused questions (e.g., "Who would wear smaller glasses?"). Children who were in the picture book condition used a variety of mathematical language more often during free play periods over the course of the study and had greater gains in mathematics than children who only received typical mathematical instruction and a picture book with no mathematics-focused questions. Green et al. (2018) found similar results for 3- to 5-year-old children, such that scripted book reading with mathematics-related questions led to gains in numeracy skills.

Some interventions involving broadly defined mathematical language have demonstrated mixed results. For example, Powell and Driver (2015) incorporated mathematical vocabulary questions and instruction (e.g., “What does the equal sign mean?”) into a tutoring program for first-grade students’ addition problem solving. Children who received this extra vocabulary instruction performed similarly to children who only received addition instruction on measures of mathematical vocabulary terms and a measure of addition fluency. Similarly, Hassinger-Das et al. (2015) found that incorporating a number sense intervention into a mathematically focused picture book that used mathematical vocabulary (e.g., same, enough, add, subtract) led to greater gains in kindergarten children’s understanding of mathematical language compared to a control group, but not broader mathematics skills. Children who were assigned to a number sense intervention without the added picture book component outperformed the children who received the picture book intervention on a measure of calculation skill at posttest. However, on other measures of mathematics achievement, the experimental conditions performed similarly to each other and to a control condition.

Limitations of Prior Work and Need for Subsequent Research

In the Appendix, we provide a list of previous intervention studies evaluating the impact of mathematical language and exact numeracy interventions (both alone and combined) with their respective impacts on mathematical language and numeracy performance across all measures and contrasts made in each study. Given the breadth of studies (both in content and design) that focus on numeracy instruction, the studies included in this table are limited to those most similar to the interventions targeting mathematical language (i.e., either those embedded in studies with a mathematical language intervention or those using picture books as the mode of intervention). No studies have explicitly contrasted a mathematical language only intervention (i.e., an intervention that does not include exact numeracy) and an exact numeracy only intervention, nor have any studies contrasted a combined condition with a quantitative mathematical language only condition. In Table 1, effect size (Hedges’ g) averages for contrasts are presented for the prior studies. These effect size averages indicate that, in contrast to control conditions, mathematical language and numeracy interventions alone had moderate effects (Hedges’ $g = 0.46$ and 0.54, respectively) within domain. However, only mathematical language (Hedges’ $g = 0.27$), and not numeracy instruction (Hedges’ $g = 0.01$), had cross-domain impacts. Moreover, as expected, mathematical language combined with numeracy instruction had benefits for both mathematical language skills (Hedges’ $g = 0.40$) and numeracy skills (Hedges’ $g = 0.45$) when contrasted with a control condition. Although no studies have directly contrasted either a numeracy intervention alone or a combined intervention with mathematical language instruction alone, there have been studies (Hassinger-Das et al., 2015; Powell & Driver, 2015) that have contrasted a combined condition with numeracy instruction alone. However, these findings were not in line with expectations that combined mathematical language and numeracy instruction should enhance numeracy performance above numeracy instruction alone (e.g., Chard et al., 2008; Clements & Sarama, 2011).

The discrepancy in findings versus expectations may be due to a number of design features of prior work. For example, the content covered in the previous studies may have influenced the effectiveness of the mathematical language instruction for supporting numeracy instruction. Powell and Driver (2015) and Hassinger-Das et al. (2015) largely focused on more formal mathematics skills (e.g., addition and subtraction). Certain numeracy skills are more language-based skills (e.g., counting, numerical comparison, story problems; Hornburg et al., 2018), and mathematical language instruction may be more effective when targeted at these types of skills. Relatedly, the outcome measures that showed an advantage for numeracy only instruction in both of these studies were children’s calculation skills (Hassinger-Das et al., 2015) or children’s addition fluency (Powell & Driver, 2015). This is in line with work showing that mathematical language is less strongly related to formal computation skills than it is to early number knowledge (Hornburg et al., 2018). Moreover, improving targeted aspects of mathematical language (i.e., quantitative mathematical language) may be more impactful for improving numeracy skills (Purpura et al., 2019). This may explain why previous work using mathematical language broadly defined has shown mixed results in improving children’s numeracy skills (e.g., Powell & Driver, 2015 focused on defining the symbols in addition problems, Jennings et al. (1992) used a wide range of mathematical terms ranging from quantitative and spatial to more applied terms relating to money and time, and Hassinger-Das et al. (2015) used a breadth of terms that included both quantitative and spatial language). As indicated by Boonen et al. (2011), too broad of a focus on mathematical language may not be supportive of children’s numeracy development as it may distract or add confusion to the acquisition of numeracy knowledge.

At present, the unique and combined effects of mathematical language and numeracy instruction on children’s early numeracy development remain unclear. The studies outlined above provide evidence that mathematical language knowledge is malleable (e.g., Hassinger-Das et al., 2015; Jennings et al., 1992; Powell & Driver, 2015) and that mathematical language instruction alone can improve children’s numeracy skills (Purpura et al., 2021; Purpura, Napoli, et al., 2017). However, given the recommendations to integrate both mathematical language and numeracy in instruction (Chard et al., 2008; Clements & Sarama, 2011), it is necessary to evaluate if there are synergistic effects of instruction on both mathematical language and numeracy skills. Although prior work has examined mathematical language instruction alone, or mathematical language combined with numeracy instruction, no study has included both such conditions. Moreover, there is mixed evidence contrasting mathematical language plus numeracy instruction versus numeracy instruction alone (Hassinger-Das et al., 2015; Powell & Driver, 2015), but these studies were conducted with older children and the alignment between the mathematical language content was somewhat broad (Hassinger-Das et al., 2015) or more focused on the names for mathematical symbols (Powell & Driver, 2015). Given that there is a strong relation between quantitative mathematical language and early numeracy skills that is evident even in

Table 1
Average Hedges' g Effect Sizes Across Previous Intervention Conditions and for the Current Intervention

Contrast	Quantitative language skills		Numeracy skills	
	Prior studies	Current intervention	Prior studies	Current intervention
Quantitative language only versus control	0.46	0.14	0.27	−0.21
Numeracy only versus control	0.01	−0.03	0.54	−0.10
Combined versus control	0.40	0.21	0.45	0.07
Quantitative language only versus numeracy only	—	0.17	—	−0.12
Combined versus quantitative language only	—	0.07	—	0.28
Combined versus numeracy only	0.38	0.23	−0.24	0.17

Note. A positive effect size indicates in favor of the first listed condition and a negative effect size indicates in favor of the second listed condition.

preschool (Purpura et al., 2019), there is a critical need to explicitly contrast instruction on quantitative mathematical language alone, numeracy alone, and both quantitative mathematical language and numeracy combined. Findings will not only advance empirical knowledge of the underlying developmental process of early mathematics but also have practical implications for instructional recommendations in preschool. The current study will use children's picture books as the intervention delivery mechanism for methodological reasons described below and because picture book reading is an important context of children's early learning environment in both homes and schools.

Prior Intervention Work Using Picture Books

Previous work has highlighted how picture books can be used to promote children's early numeracy skills (Gibson et al., 2020; Green et al., 2018; Jennings et al., 1992; Mix et al., 2012; O'Rear & McNeil, 2019; Purpura et al., 2021; Purpura, Napoli, et al., 2017; Young-Loveridge, 2004). This work has largely focused on picture books because they provide a medium where the input (i.e., the text on the page or the script the experimenter follows) is easily manipulable without needing to change the structure of other aspects of the intervention. This approach to intervention work has been used to identify ways to structure number talk to promote children's understanding of cardinality (e.g., Mix et al., 2012) as well as ways to deliver mathematical language to promote children's mathematical language skills (e.g., Purpura, Napoli, et al., 2017). These interventions have the added benefit of being easily translatable to real-world circumstances. One-on-one work with trained experimenters in school settings has shown that picture books are an effective way to promote children's understanding of cardinality in children from middle-SES homes (Mix et al., 2012) as well as children attending Head Start programs (O'Rear & McNeil, 2019). Previous work has also shown that picture books can be effective ways to promote numeracy skills when implemented by educators at the whole-class level (Jennings et al., 1992) as well as at the one-on-one level by school staff (Young-Loveridge, 2004). At home, picture books provided to parents are effective ways to promote mathematical language and broader numeracy skills (Purpura et al., 2021) as well as more targeted numeracy skills such as an understanding of cardinality (Gibson et al., 2020). Given the ease of incorporation of picture books both at home and in the classroom, and because using picture books would allow us to provide consistent delivery across conditions while only manipulating the target constructs, picture books were chosen as the delivery mechanism for our intervention.

Current Study

The goal of the current study was to examine the unique and combined influence of both quantitative mathematical language and exact numeracy instruction on children's early numeracy skills within the context of shared book reading. Specifically, the effects of four intervention conditions (quantitative language only, numeracy only, a combination of both quantitative language and numeracy instruction, and an active control group) on preschool children's quantitative mathematical language and numeracy skills were contrasted. The interventions used picture books to deliver the content of the intervention to both answer the basic research question and to lead to a practical intervention that can later be used in schools (e.g., Gibson et al., 2020; Mix et al., 2012; O'Rear & McNeil, 2019; Purpura et al., 2021; Purpura, Napoli, et al., 2017). The quantitative language only condition targeted children's understanding of quantitative mathematical language terms (e.g., more, many, few, fewest, same, different), but explicitly avoided exact numeracy (e.g., counting, labeling exact set size, addition). The exact numeracy only condition included exact numeracy instruction previously shown to promote children's early numeracy understanding but did not include any quantitative mathematical language. The combined condition included both quantitative mathematical language and exact numeracy instruction. These conditions were contrasted with an active control condition that used books that do not include any quantitative mathematical language or numeracy. It was hypothesized that:

Hypothesis 1 (H1): Children who participated in the quantitative language only condition would outperform children in the active control condition on quantitative mathematical language (H1a) and numeracy skills (H1b).

Hypothesis 2 (H2): Children who participated in the numeracy only condition would outperform children in the active control condition on numeracy skills (H2a). We did not expect the numeracy only condition to impact quantitative mathematical language as the average prior effect sizes are <0.10 (H2b).

Hypothesis 3 (H3): Children who participated in the combined [quantitative language + numeracy] condition would outperform children in the numeracy only condition on both quantitative mathematical language (H3a) and numeracy skills (H3b), as well as outperform children in the quantitative language only condition on numeracy skills (H3c).

We did not hypothesize that performance between the combined and quantitative language only conditions would differ on quantitative mathematical language. H1 and H2 (positive effects of each type of instruction on its own) were informed by previous research showing that quantitative mathematical language instruction improves both quantitative mathematical language and numeracy skills (Purpura et al., 2021; Purpura, Napoli, et al., 2017) and that exact numeracy instruction from shared book reading is an effective way to promote numeracy skills (Gibson et al., 2020; Mix et al., 2012; O'Rear & McNeil, 2019). H3 (an additive effect of quantitative mathematical language instruction plus numeracy instruction) was informed by the idea that quantitative mathematical language instruction provides children with a broader approximate understanding of the relations between quantities which in turn helps children learn the more exact concepts underlying numeracy skills from the numeracy instruction (Purpura et al., 2019, 2021), but that the combination of numeracy and quantitative mathematical language instruction should not enhance the effects of the intervention on quantitative mathematical language development. Such findings would indicate that quantitative mathematical language underlies and enhances numeracy development.

Method

Participants

Proposed sample: It was expected that at least 220 participants would be recruited for this study based on a power analysis conducted in PowerUp! (Dong & Maynard, 2013) using effect sizes from prior studies (details on the power analysis procedures and assumptions can be found later in the "Method" section in the "Power Analysis" subsection). We planned to recruit children ages 3–5 years old from preschools across the state of Indiana, United States, that primarily serve children from families with low incomes. Participants in previous studies that were recruited from similar preschool populations were 14% Latine, 34% Black, 45% White, and 7% other or multiracial. Therefore, we expected the final sample for the current study to be racially and ethnically diverse. This study has been approved by the Purdue University Institutional Review Board.

Actual participants: A total of 289 participants from 18 preschools were recruited to participate in the study. Of those children, 55 withdrew or did not meet eligibility requirements before being randomly assigned to conditions. Reasons for withdrawal included: moving ($n = 14$), limited availability during the school day (e.g., only attended 2 days per week; $n = 14$), had a previously unreported developmental delay ($n = 14$), spoke limited English ($n = 8$), or did not provide assent ($n = 5$). As a result, 234 children were randomly assigned to one of the four conditions: quantitative language only ($n = 60$), numeracy only ($n = 57$), combined ($n = 58$), or active control ($n = 59$). The 234 children who were assigned to condition ranged in age from 3.03 to 5.61 years old ($M = 4.29$, $SD = 0.62$). Overall, 54% of participants were female. For parental education, 38% had less than a college degree, 31.2% had a 2- or 4-year college degree, 28.3% had a graduate degree, and 3.4% did not report parental education. Race/ethnicity for participants was 65% White, 12% multiracial, 9.8% Black/African American, 8.1% Asian, 3.4% Latine, and 1.7% not reported.

Among children who were assigned to the condition, 16 (6.84%) left the study before posttesting because they either left the school or their family moved. Attrition by group was as follows: quantitative language only ($n = 1$; 1.67%), numeracy only ($n = 8$; 14.04%), combined ($n = 3$; 5.17%), and active control ($n = 4$; 6.78%). There was no substantial differential attrition (What Works Clearinghouse, 2022) between groups for the key contrasts.

Measures

After consent from parents was obtained, trained assessors worked with children. The same achievement measures were used at both pretest and posttest. Assessments took place in the participating preschools in an area designated by the preschool teachers or directors. Assessors were either paid project staff who had completed or were working toward completion of their bachelor's degree or undergraduate students working on the project for academic credit. All assessors completed extensive training provided by the lead project staff and were blind to the intervention condition. This training included individual introduction and practice sessions for each measure, opportunities to practice on the measures, and a "testing out" session that required accurate testing procedures while implementing the measures. Although it was planned to do the testing out session with actual children, this was not feasible due to constraints in school access as a result of the COVID-19 pandemic. Thus, the lead project staff conducted mock settings where they simulated a testing session while pretending to be a preschool child.

The direct child measures included measures of numeracy skills, quantitative mathematical language skills, general vocabulary skills, and cognitive flexibility. Testing on these measures at each time point was approximately 20–40 min per child and was broken into smaller testing sessions as needed so as not to overtax the children. Individual pretest sessions took place during the fall semester of the school year (October to December 2021). The intervention was planned to take place across an 8-week period in the spring semester (mid-January to early March 2022) but was extended through April due to COVID-19 closures for some classrooms, and all posttest sessions occurred after the intervention was completed in individual schools (mid-March to May 2022).

Parent Survey

Parents were asked to complete a basic demographic survey that included their child's birthdate, gender, race/ethnicity, highest parental education, and identification of any diagnosed developmental delays.

Numeracy Skills

The Preschool Early Numeracy Screener (PENS) is a standardized, norm-referenced 25-item test of numeracy skills (Purpura, 2021). The items are representative of the broad numeracy skills children are expected to attain in preschool and assess key domains identified as being critical precursors of mastery in mathematics. Test items cover topics such as verbal counting, exact comparisons, one-to-one correspondence, number order, numeral identification, ordinality, and number combinations. For example, a one-to-one correspondence item may show the child a picture of three dots, and the research assistant says, "Count these dots." The test has high internal consistency ($\alpha = .89$) and is highly related to other measures of numeracy ($rs > .80$; Purpura, 2021). Children's total scores were computed as the sum of their total correct responses.

Although three items on this measure include a quantitative mathematical language word (most), it was not expected that the inclusion of those items would impact the findings as most is one of the easiest mathematical language terms (Purpura & Reid, 2016). However, to add a robustness check to the analyses, the results were also examined and reported both with and without those items factored into the total score.

Quantitative Mathematical Language

A modified version of the Preschool Assessment of the Language of Mathematics (PALM; Purpura & Logan, 2015) that is exclusively focused on quantitative mathematical language and expands the measurement of quantitative mathematical language beyond the original measure was used for this study. The quantitative mathematical language measure included 14 items (more, most, fewest, fewer, least, less, a lot, a little bit, add, take away, same [quantity], same [visually], similar, different; $\alpha > .76$). All items were designed to be completed without exact quantitative skills or in a nonnumeracy context. Scores were computed as the sum of children's correct responses.

General Vocabulary

General vocabulary was included as a covariate because it has been shown to be related to, and potentially be a precursor of, children's mathematical language skills (Toll & Van Luit, 2014). Children's receptive vocabulary was assessed using the National Institutes of Health (NIH) Toolbox Picture Vocabulary test (Weintraub et al., 2013). Theta scores provided by the NIH Toolbox output were used in analyses. This assessment is considered to have strong discriminant and convergent validity in children ages 3–15, as well as strong test–retest reliability intraclass correlation coefficient (ICC = .84; 95% CI [0.75, 0.90]; Weintraub et al., 2013).

Cognitive Flexibility

Cognitive flexibility was used as a covariate to represent executive functioning skills, which are robustly linked to numeracy development (Allan et al., 2014). This measure of cognitive flexibility is a good proxy for children's attention skills during reading, which is an important covariate to include given that the intervention will be implemented in small groups. Cognitive flexibility, also called attentional flexibility, reflects one's ability to focus and sustain attention voluntarily while adapting or shifting attention when necessary (e.g., to changing stimuli; Rothbart & Posner, 2005). A card sorting task that is based on the dimensional change card sort (DCCS; Zelazo, 2006) was used to measure children's cognitive flexibility. On this measure, children were asked to sort picture cards on the basis of three different dimensions: shape, color, and size. This task consisted of three mandatory sections consisting of six items each. For the first section, children were asked to sort on the basis of shape. For the second section, the rule changed and children were asked to sort on the basis of color. For the third section, children were asked to sort on the basis of size. If children scored 5 or more points on the third section, a fourth set of six items was administered which consisted of a more complex rule that required children to sort on the basis of size when a card included a thick black border and to sort on the basis of color when the card did not have a thick black border. One point was given for each correct response, with scores ranging from 0 to 24. This measure has shown strong psychometric properties in previous research ($\alpha > .80$; McClelland et al., 2014; Schmitt et al., 2018).

Procedure

Overview and Proposed Timeline

Recruitment for participants began in August 2021 and was expected to continue through October 2021, or until we reached our necessary sample size of 220 children. However, given the ongoing concern with potentially losing schools/classrooms to closures due to teacher shortages, we decided to recruit more children at collaborating schools and reached a total of 289 consented children. Pretesting began in October 2021 and continued through December 2021. Random assignment to condition occurred after winter break in January 2022 so that children who did not return to school were not assigned to condition. Children were block-randomly assigned within classroom to one of four conditions: (a) quantitative language only, (b) numeracy only, (c) combined, and (d) active control. When classrooms had fewer than four participating children, we combined classrooms for blocking, prioritizing classrooms within the same school whenever possible. The 8-week interventions began in January 2022 and were intended to continue through early March 2022. However, due to the COVID-19 pandemic, specifically the Omicron Wave, participating schools/classrooms had several instances of 2-week closures for quarantines. When these instances happened, the intervention implementation in those schools or classrooms was delayed for the duration of the closure and resumed where it had left off when they reopened. This resulted in some of the schools continuing the intervention into April 2022. Fortunately, because assignment to condition happened within the school, the impacts of these delays were equally held across conditions. Additionally, once schools discontinued the practice of entire school quarantines, some children were out of school for individual child/family quarantines. As a result, and to make practices consistent across schools, we elected to make-up sessions for children on individual/family quarantines within 2 weeks of their return to school when possible. Moreover, no more than three sessions were made up in any week so that children were not overburdened by extra sessions. Children participated in intervention sessions in small groups (two to four children) three times per week for an intended total of 8 weeks (max of 24 sessions). Posttest assessments began in March 2022 and continued through May 2022.

Intervention Process

Intervention Design. Trained project staff conducted the intervention sessions. These project staff were graduate students, postdoctoral researchers, and other paid project staff with a completed bachelor's degree or extensive experience working with young children. After children were assigned to a condition, they engaged in their intervention as prescribed for 3 days/week for 8 weeks in small groups (two to four children). Across all conditions, there were a total of 80 small groups (20 quantitative language only; 20 numeracy only; 21 combined; 19 active control). Each intervention session took approximately 15 min. Small group instruction, rather than individual instruction, was selected to reduce classroom interruption, reduce overall instruction time for project staff, and make the intervention more feasible to complete. Our original intent for student absences was to follow the procedure that when individual children were absent on a day of intervention instruction, interventionists would attempt to schedule an individual session with the absent child within 1 week of their return to school and interventionists would only schedule make-up sessions for absent children within 1 week of their absence. However, due to the issues with COVID-related absences, this window for make-up sessions was extended to 2 weeks for individual absences. When schools or classrooms were closed due to quarantines, the intervention was paused and then resumed where it left off when the school reopened. Attendance logs were kept for each child to calculate a child's total participation in the intervention. Attendance (number of completed sessions) was included in analyses as a covariate.

Intervention Conditions. The four conditions all followed a similar structure. This structure included three picture book reading intervention sessions per week for 8 weeks. Each condition had three books that explicitly targeted the intended content (quantitative language only, numeracy only, combined, or neither [active control]). The quantitative language only, numeracy only, and combined conditions all used modified versions of the same three books to ensure that only the explicit instructional content differed. The comparison books (i.e., the active control) were designed to not include any quantitative mathematical language or numeracy content in the text, dialogic reading prompts, or illustrations.

Intervention Books. Children in the three mathematical intervention conditions were read the same three picture books from the *Little Elephants' Big Adventures* series. These three picture books are titled *Too Many Pillows*, *Just Enough Eggs*, and *Picnic with Some Peanuts*, and were all written by Angela M. Isaacs and illustrated by Matt Dye. These three picture books were designed to increase the diversity of the input children received across the intervention. That is, each book introduced new target content that previous books had not covered. In addition to the story text, these books have three dialogic reading prompts per page designed to engage children in the story and to highlight the focal content. The dialogic reading prompts were color-coded so that the interventionist used one prompt per reading and the prompts became more complex with each reading. The books were originally developed to focus on quantitative language (Purpura et al., 2021); however, the books were modified so that there were three separate versions of each book: The original version with only quantitative mathematical language, a version with only exact numeracy, and a version with both quantitative mathematical language and exact numeracy (combined). An example of the text and dialogic reading prompt differences across the versions can be found in Table 2.

Quantitative Language Only Condition. The quantitative language picture books included target words within the text as well as dialogic reading questions to promote children's engagement with mathematical language. To isolate the effects of quantitative mathematical language, there was no mention of counting, cardinality, or addition within the text. Each of the first 3 weeks focuses on one book each week before the same 3-week schedule is followed for Weeks 4 through 6 (see Table 3). Across the books, new quantitative mathematical language was introduced (e.g., *Too Many Pillows* target words include a lot, many, more, and most, *Just Enough Eggs* target words include same, similar, and different, and *Picnic with Some Peanuts* target words include few, fewest, and some). Thus, across the weeks, the books built in complexity by focusing on expanding children's understanding of mathematical language.

Numeracy Only Condition. The numeracy only condition used the same books as the quantitative language only condition, but the text and questions were modified to remove any quantitative mathematical language and were replaced with counting, cardinality, and addition content. To mirror the quantitative language only condition's progress over the weeks, children in the numeracy only condition read books that increased in difficulty across the 3 weeks. In *Too Many Pillows*, children primarily focused on smaller set sizes (1–3) while reading the book and were asked to both count and label the sets. In *Just Enough Eggs*, children were primarily focused on larger sets (>3) and were asked to both count and label the sets. In *Picnic with Some Peanuts*, children continued to focus on the cardinality of the sets and there were dialogic reading questions relating to arithmetic (e.g., "If Lucy ate four of her apples, how many apples would she have left?").

Combined Condition. The combined condition included content from both the quantitative language only condition and the numeracy only condition. For example, in Week 1 of the combined condition, children read a book that included the same focus on mathematical language as the quantitative language only condition (i.e., *Too Many Pillows* target words include a lot, many, more, and most) and the same focus on set size as the numeracy only condition (i.e., the focus was primarily on sets 1–3 with counting and labeling of the set size). To control for the overall level of math-related input, the books were designed so that there was a roughly equal balance in terms of the focus on quantitative mathematical language and exact numeracy. That is, when adapting these books from the originals (those used in the quantitative language only condition), approximately half of the mathematical language terms and dialogic reading questions were replaced with a focus on exact numeracy.

Active Control Books. To provide a strong counterfactual to the intervention groups, an active control group was used. The books that were used in this group were also researcher-designed in a similar format to the *Little Elephants*' books, including length and structure. However, the three books used for the active control condition did not include the use of quantitative mathematical language or numeracy content. These three books are titled *Forgetful Fred Gets Ready for Bed*, *Maria's Perfect Day*, and *Bella Gets Ready to Ride* and were all written by Angela M. Isaacs and illustrated by Matt Dye. These books also included three color-coded dialogic reading prompts on each page to engage children more in the story. These books have been used in an active control condition in prior work (Purpura et al., 2021).

Intervention Structure. An overview of the intervention structure is presented in Table 3. In each condition, a different book was read three times each week for the first 3 weeks. Then, each book was read one more time in Week 4 using the third prompt in each book to reinforce the previous readings. The same process was then repeated for Weeks 5 through 8. This structure was intended to enable children to consolidate information from the first 6 weeks.

Project staff read the book to a small group of children. During the reading session, project staff read one dialogic reading question in the appropriate color prompt each time they read a page in the book assigned that week. Interventionists scaffolded dialogic reading responses using the PEER framework (Prompt, Evaluate, Expand, Repeat; Whitehurst, 1992). If children responded correctly, the project staff provided reinforcement of the response by repeating the correct response (e.g., "Yes, she does have more.")

Table 2
Examples and Modifications to the Text and Dialogic Reading Questions Across Conditions

	Quantitative language only	Numeracy only	Combined
Text	"Whoops!" Bear fell down. "Bear," scolded Lucy. "Now the sugar is a different amount." Benjamin got more sugar. "Now they're equal amounts again."	"Whoops!" Bear fell down. "Bear," scolded Lucy. "You knocked over two cups of sugar." Benjamin got two new cups of sugar. "Now they're equal amounts again."	"Whoops!" Bear fell down. "Bear," scolded Lucy. "Now the sugar is a different amount." Benjamin got two new cups of sugar. "Now they're equal amounts again."
Question 1	1. Did Benjamin get enough sugar?	1. Do Benjamin and Lucy have four cups of sugar now?	1. Do Benjamin and Lucy have four cups of sugar now?
Question 2	2. Why did Benjamin get more sugar?	2. Why did Benjamin get two new cups of sugar?	2. Why did Benjamin get two new cups of sugar?
Question 3	3. Why is Lucy worried that some of the sugar spilled?	3. Do Lucy and Benjamin have the four cups of sugar they need? Can you count them?	3. Why is Lucy worried that some of the sugar spilled?

and then continued to the next page. If children responded incorrectly, the project staff evaluated the response and expanded and corrected that response in a constructive manner. Interventionists were explicitly trained to use only the correct terminology (e.g., mathematical language, numeracy) in their respective conditions. The questions were posed generally to the group and project staff worked to ensure comparable opportunities for answering across all children.

Fidelity. Interventionists audio-recorded each intervention session and one third of each interventionists' audio files were randomly selected to be transcribed and coded. To assess intervention fidelity, interventionists were evaluated on whether or not they used each of the prespecified dialogic reading prompts. Fidelity rate was computed as the percentage of correctly used prompts throughout a storybook reading session. Each book contained 14 total prompts to be read each session. Therefore, fidelity out of the total number of prompts used correctly was averaged out of a 14-point score.

Of the 2,381 total audio files completed, one-third ($n = 794$) were randomly selected to be assessed for fidelity. An additional 20% ($n = 157$) of the randomly selected 794 files were double-coded. Interrater reliability ranged from 0.75 to 1.00 and had an average score of 0.97. Overall, fidelity of the intervention was high. Across the 15 interventionists, average fidelity was 97% with a range of 83%–100%. Fidelity was consistently high across conditions: quantitative language only ($n = 210$; 97%), numeracy only ($n = 201$; 100%), combined ($n = 204$; 95.5%), and active control ($n = 179$; 99%).

We planned to conduct weekly interventionist fidelity checks to ensure the intervention was delivered correctly. If an interventionist's fidelity rate was below 80% (i.e., fewer than an average of 12 out of 14 prompts per book correct), they would have been provided with additional training within the next week. We expected that interventionists would be able to deliver these interventions with high fidelity because in prior work using a similar style of reading prompts, caregivers implemented the intervention with an average fidelity rate of 91% with minimal training and no follow-up (Purpura et al., 2021). As part of these fidelity checks at the end of each week, a random third of the audio-recorded sessions for each interventionist was reviewed to identify if the interventionist used noncondition terminology (e.g., use of quantitative mathematical language in the numeracy instruction only condition). These weekly checks were used to maintain treatment integrity. If noncondition terminology was observed, the interventionist would be provided with additional training and support to ensure treatment fidelity. After the first 3 weeks, these weekly fidelity checks were discontinued because interventionists demonstrated high fidelity. Moreover, due to the impacts of the pandemic, personnel resources were needed on other aspects of the project (e.g., scheduling, participant tracking, make-up sessions).

Table 3
Intervention Delivery Structure

Condition book	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Quantitative only	QB1×3	QB1×3	QB2×3	QB1×1 QB2×1 QB3×1	QB1×3	QB2×3	QB3×3	QB1×1 QB2×1 QB3×1
Numeracy only	NB1×3	NB2×3	NB3×3	NB1×1 NB2×1 NB3×1	NB1×3	NB2×3	NB3×3	NB1×1 NB2×1 NB3×1
Combined	QN1×3	QN2×3	QN3×3	QN1×1 QN2×1 QN3×1	QN1×3	QN2×3	QN3×3	QN1×1 QN2×1 QN3×1
Active control	NM1×3	NM2×3	NM3×3	NM1×1 NM2×1 NM3×1	NM1×3	NM2×3	NM3×3	NM1×1 NM2×1 NM3×1

Note. For each condition, book one refers to the different versions of *Too Many Pillows,* book 2 refers to the different versions of *Just Enough Eggs*, and book 3 refers to the different versions of *Picnic with Some Peanuts*. QB = quantitative language book; NB = exact numeracy book; QN = combined [quantitative language + numeracy] book; NM = nonmathematics book.

Analytic Pipeline

Inclusion and Exclusion Criteria

Inclusion. To be eligible to participate, children must have (a) attended one of the 13 designated preschools, (b) had parental consent to participate, (c) been 3–5 years old, (d) been English-speaking, (e) had no known developmental disorders, and (f) given assent to participate. The children who met these six criteria were included in the study.

Exclusion. All children who were assigned to a condition were included in the final data analysis as an intent-to-treat model was used. Even if children assigned to a condition did not participate in the intervention sessions, but had pretest data, they were included. No outlier protocol was used for children's performance as it was assumed that any outliers were representative of a child's actual ability. Thus, in all cases in which children completed pretesting and were assigned to condition, they were included in analyses.

Analytic Plan

In this study, we randomly assigned individual children to one of four conditions in a 2 (numeracy) × 2 (quantitative mathematical language) design (quantitative language only, numeracy only, combined, active control). Therefore, for any specific statistical contrast, we considered 25% of the sample to be in the treatment condition.

Large-scale nationally representative datasets indicate that conditional ICCs for math are low in early elementary school (ICC < .11; Hedges & Hedberg, 2007) and similarly designed intervention studies have indicated that classroom effects were negligible (Hassinger-Das et al., 2015; Purpura, Napoli, et al., 2017). However, we planned that regardless of whether or not the ICCs at intervention group, classroom, or school were negligible, we would adjust the *t* test and degrees of freedom for estimates of statistical significance post hoc, using standard error adjustment procedures described by Hedges (2007) to correct for clustering. Ultimately, we found that ICCs for the quantitative language outcome was negligible (<.001) and any adjustments would not have changed *p* values. For the numeracy outcome, the ICC was .11 (.095 for the modified version used in the robustness check). Because adjusting nonsignificant *p* values would not change interpretations, only the findings with *p* values <.100 were adjusted. Moreover, robustness check analyses were conducted as three-level multilevel models to account for variance at both the group and classroom levels. Differences between that method and the planned analyses were negligible, thus we retained our planned and more parsimonious models where we adjusted *t* test and degrees of freedom for estimates of statistical significance. Primary analyses followed a regression-based framework with three dummy-coded variables representing the quantitative language only, numeracy only, and combined conditions for H1 and H2. An equation representing this analytic model is:

$$\begin{aligned} Y = B0 + B1 \times (\text{Numeracy}) + B2 \\ \times (\text{Quantitative Mathematical Language}) + B3 \\ \times (\text{Combined}) + B4 \times (\text{Covariates}). \end{aligned} \quad (1)$$

The covariates term in the equation is a shorthand representation of an array of all covariates used in the analysis (covariates are listed in the next section). The analysis was repeated separately for each posttest measure (Y) and, following a residualized gain format, we also controlled for all covariates including the corresponding pretest measure. For H3, the reference group was changed from the active control group to the combined group, but all other analytic procedures remained the same.

For H1a, we interpreted coefficient B2, as it represents the difference between the quantitative language only condition and the active control condition, for the quantitative mathematical language outcome (Y). For H1b, we examined the same coefficient when numeracy was the outcome variable (Y). For H2a, we again examined the numeracy outcome, but focused on the statistical significance of coefficient B1 (representing the difference between those children assigned to the numeracy only condition and the active control condition). For H2b, we examined the same coefficient when quantitative mathematical language was the outcome variable. H3 contrasted the combined condition with those who received the numeracy only intervention on their quantitative mathematical language (H3a) and numeracy (H3b) skills. We also compared children in the combined group with children in the quantitative language only group on their quantitative mathematical language (H3c). H3a–H3c were tested using post hoc contrasts with a conservative Bonferroni correction to the critical *p* value (α level = .025).

Covariates

In addition to pretest quantitative mathematical language and numeracy performance, analyses included the following covariates: child age, gender, parental education, general vocabulary, cognitive flexibility, and attendance (i.e., the number of intervention sessions in which the child participated). Group differences on pretest variables were tested by conducting analyses of variance (ANOVA) with the key pretest scores as the dependent variable and the conditions as the independent variables. These results are presented in the preliminary analyses section of the "Results" along with descriptive data on the measures and a correlation table.

Missing Data

It was expected that in some cases, it may have been possible that children were missing data at random. To counter this loss of information, full implementation maximum likelihood estimation was intended to be used in all primary analyses. Maximum likelihood is an efficient method for handling most types of missing data but does not support missingness on dichotomous predictors or on the dependent variable (cases are deleted listwise when there is missingness on these). Therefore, we planned that if we were unable to conduct maximum likelihood estimation due to the nature of the missingness, we would instead use multiple imputation methods. Multiple imputation was planned to be conducted following steps to evaluate the nature of the missingness and attrition as outlined by the What Works Clearinghouse (2020). Additionally, we planned to follow the same procedures as explained for testing potential pretest differences to test for covariates and other variables that may be related to missingness on all outcome variables. Anything that was found to be related to the missingness (an effect size > 0.25) was included as an auxiliary variable in the imputation model. We planned to impute 10 datasets and pool results across imputations. Ultimately, all participants who were randomly assigned to the condition had complete assessment data at pretest. In terms of demographic data included in the analyses, eight of the 239 participants were missing parent education data. The only missing data at posttest was due to attrition. Thus, because the primary missing data were on posttest measures, multiple imputation with 25 imputed datasets with results pooled across imputations was used. The modified number of 25 imputed datasets, rather than the proposed 10, was used due to estimation convergence. The imputation model included all covariates (child age, gender, parent education, and attendance), as well as all pre- and posttest measures.

Power Analysis

It was expected that a total sample of 200 participants (50 per condition) would be needed to achieve adequate power. This a priori power analysis was conducted using the Hedges' *g* effect size averages (see Table 1) from the systematic review of the mathematical language and numeracy intervention literature (see the Appendix). From the identified studies, effect sizes were harvested, and a single Hedges' *g* was estimated for each hypothesized contrast and each outcome. The expected effect size of the intervention varied depending on the conditions being contrasted and the outcome (mathematical language or numeracy). Given that 200 participants were needed, we aimed for an additional 10% to account for attrition that was expected to occur during the course of the intervention. Thus, we aimed to recruit a total of at least 220 participants (55 per condition).

Minimal Detectable Effect Sizes (MDESs)

In previous work, when mathematical language was the outcome (note that some of these studies used a combined measure of quantitative and spatial mathematical language), the effects of similar interventions were generally moderate (Hedges's *g*: 0.38, 0.40, and 0.45). For the numeracy outcome, the smallest effect sizes from the published literature were those contrasting quantitative language only intervention groups versus Control (effect size of 0.27 in our review), and numeracy only versus combined (numeracy and quantitative language) conditions (effect size of −0.25). Although the effect size in the prior literature was negative (favoring numeracy only instruction), it was our expectation that, because of the design of our intervention (tightly connected quantitative language and numeracy content and within a preschool population), the combined condition would outperform the numeracy only condition. Therefore, we used the smallest MDES of 0.25 to determine the sample size necessary for the present study because by using this estimate, all other primary analyses were sufficiently powered.

More specifically, for H1, we expected that children who participated in the quantitative language only intervention condition would outperform children in the active control condition on quantitative mathematical language (H1a; expected effect size = 0.45) and numeracy (H1b; expected effect size = 0.27). For H2, we expected that children who participated in the numeracy only condition would outperform children in the active control condition on their numeracy skills (H2a; expected effect size = 0.54), but not on quantitative mathematical language (H2b; expected effect size = 0.01). For H3, we expected that children who participated in the combined condition would outperform children who received the numeracy only condition on quantitative mathematical language (H3a; expected effect size = 0.38) and on their numeracy skills (H3b; expected effect size = 0.25). For our final hypothesis, we expected that children who received the combined condition would also outperform children in the quantitative language only condition on numeracy (H3c; expected effect size = 0.25), but not on quantitative mathematical language.

Power Calculation. As previously described, to determine the sample size necessary for this model, we used PowerUp! (Dong & Maynard, 2013), with two-level blocked random assignment (using the module "fixed effects and interactions with treatment"). The MDES for numeracy outcomes was smallest and was used to power the study (Hedges's $g = 0.25$). We used an α of .025 (for the anticipated post hoc contrasts), a power estimate of 0.90, and relied on one-tailed statistical tests (because we only have directional hypotheses). Further, we relied on $R^2 = .72$ for covariates based on recent work (Purpura et al., 2021). We estimated that 200 participants (blocked within classroom) would be needed to have a power of 0.90 to detect a significant difference of an effect size of 0.25. We planned to recruit an additional 10% ($n = 20$) to account for potential attrition for a total sample of 220 (55 per condition). During our recruitment process, we chose to over recruit for our sample because we had concerns about schools/classrooms potentially closing due to teacher shortages. Although we planned to test for nesting effects (ICCs) and to adjust the *t* test and degrees of freedom for estimates of statistical significance post hoc using procedures described by Hedges (2007) regardless of ICCs, we did not incorporate nesting effects into our power analysis because prior similar studies (Hassinger-Das et al., 2015; Purpura, Napoli, et al., 2017) have found them to be negligible (e.g., Purpura, Napoli, et al., 2017 found conditional ICCs of <0.001). Second, we incorporated design features (random assignment of individual students within the classroom) that randomly distributed classroom effects across conditions. We also had fairly scripted intervention procedures using books and dialogic reading questions that were likely to lead to low levels of treatment heterogeneity across intervention groups. Fourth, because the research questions were about the differences between randomly assigned children and did not involve any features of classrooms, multilevel approaches were not necessary (McNeish et al., 2017). Moreover, after reviewing the prior literature, this study would use larger sample sizes per condition than any of the other studies included in this power analysis (see the Appendix).

Transparency and Openness

This study was originally accepted as a Stage 1 in-principle acceptance (IPA) at the Journal of Educational Psychology and the accepted version, Stage 1 IPA documentation, and preregistration can be found here (https://osf.io/dbx7c/). A deidentified version of the analytic data set can be found here (https://osf.io/vbwkf/) and the code and results for primary analyses have been posted here (https://osf.io/9vrju/). All measures that were used in this study have been cited in the Method section and references are included in the reference section. The books used in the intervention study for the quantitative language only condition are available on Amazon, but the other two conditions are not yet publicly available. No substantial changes to the design and implementation of the study were made with the exceptions noted earlier to address COVID-19-related implementations. All conducted analyses follow the planned and approved proposed analyses.

Results

Preliminary Analyses

ANOVAs were conducted on all pretest variables (quantitative language, numeracy, general vocabulary, cognitive flexibility, age, parental education, gender). No significant differences among the groups were observed ($ps = .056–.950$). Correlations between key variables are presented in Table 4 and descriptive statistics are presented in Table 5. Intraclass correlation estimates from models including only the covariates showed zero to a very small proportion of variance was attributable to the classroom (0% for quantitative math language skills and 11% for numeracy skills).

Primary Results

The results for all hypotheses are presented in Table 6 and the condition variable was dummy-coded. For H1 and H2 the variables representing the assigned condition (quantitative language only condition, numeracy only condition, and combined condition) are a test of the contrast between that given condition and the control condition. For H3, the condition variable was recoded to contrast the conditions to the combined condition. For contrasts with a noted directional hypothesis, one-tailed p values were used. All other p values are two-tailed.

Hypothesis 1 (H1)

For H1, we hypothesized that children who participated in the quantitative language only condition would outperform children in the active control condition on quantitative mathematical language (H1a) and numeracy skills (H1b). However, there was no significant advantage for children in the quantitative language only condition (compared to active control) for either the quantitative mathematical language outcome (H1a: estimate $= 0.32$, $p = .148$ [one-tailed], Hedges' $g = 0.14$) or for the numeracy outcome (H1b: estimate $= -1.35$, $p = .970$ [one-tailed], Hedges' $g = -0.21$).

Hypothesis 2 (H2)

For H2a, we expected that children who participated in the numeracy only condition would outperform children in the active control condition on numeracy skills. We found that there was no difference between the numeracy treatment and control group on numeracy skills (H2a: estimate $= -0.62$, $p = .793$ [one-tailed], Hedges' $g = -0.10$). For H2b, we did not expect the numeracy only condition to impact quantitative mathematical language. Similar to previous results, the numeracy only condition did not impact quantitative mathematical language (H2b: estimate $= -0.06$, $p = .850$ [two-tailed], Hedges' $g = -0.03$).

Hypothesis 3 (H3)

To test H3, we changed the reference group from the active control group to the combined (math language + numeracy) condition—note here that negative estimates are in favor of the combined condition. We hypothesized that the combined condition would outperform children in the numeracy only condition on both quantitative mathematical language (H3a) and numeracy skills (H3b), as well as outperform children in the quantitative language only condition on numeracy skills (H3c). These contrasts provide a more direct test of the magnitude of the advantages afforded in the combined condition versus instruction in either quantitative mathematical language or numeracy alone. For H3a, we focused on the outcome of Math Language. Here we find that those in the combined condition did outperform those in the numeracy only condition on math language (H3a: estimate $= -0.53$, $p = .045$ [one-tailed], Hedges' $g = -0.23$) in favor of the combined condition. However, using the conservative Bonferroni correction, this was no longer statistically significant.

For the numeracy outcome, we find that the combined condition did not afford an advantage when compared to the numeracy only condition (H3b). In other words, the addition of math language did not improve children's numeracy skills beyond receiving numeracy alone (H3b: estimate $= -1.04$, $p = .128$ [one-tailed], Hedges' $g = -0.17$). We did, however, see that the combined condition outperformed the quantitative language only condition on numeracy skills (H3c; estimate $= -1.77$, $p = .019$ [one-tailed], Hedges' $g = -0.28$).

Robustness Check

To ensure that any potential findings on numeracy outcomes were not driven by the overlap between mathematical language and

Table 4
Correlations Between Key Variables and Covariates

Variable	1	2	3	4	5	6	7	8	9
1. Age	—								
2. Attendance	.029*	—							
3. Pretest cognitive flexibility	.430	−.013*	—						
4. Pretest vocabulary	.355	.067*	.505	—					
5. Pretest quantitative language	.441	−.017*	.547	.517	—				
6. Pretest numeracy (full)	.483	−.023*	.604	.480	.610	—			
7. Pretest numeracy (adjusted)	.463	−.032*	.594	.441	.575	.990	—		
8. Posttest math language	.431	.084*	.572	.493	.615	.614	.593	—	
9. Posttest numeracy (full)	.457	−.117**	.575	.473	.613	.770	.765	.670	—
10. Posttest numeracy (adjusted)	.458	−.107*	.572	.463	.604	.765	.763	.660	.993

Note. Numeracy (full) scores are the raw PENS scores inclusive of all items. Numeracy (adjusted) scores are the raw PENS scores minus the three items that include math language and were the scores used for the robustness check. All correlations are statistically significant ($p < .001$) except those noted. PENS = Preschool Early Numeracy Screener.
* $p \geq .010$. ** $p < .010$.

Table 5
Descriptive Statistics for Key Variables Overall and by Group

Variable	All participants				Quantitative language only condition			Numeracy only condition			Combined condition			Active control condition		
	N	*M*	*SD*	Group differences (*p*)	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Age	234	4.29	0.62	.950	60	4.27	0.60	57	4.32	0.60	58	4.26	0.63	59	4.30	0.66
Attendance	234	18.76	5.40	.080	60	19.95	3.92	57	19.14	5.94	58	18.45	5.36	59	17.47	6.01
Pretest																
Cognitive flexibility	234	10.99	5.85	.525	60	10.50	5.53	57	11.79	6.04	58	11.31	6.25	59	10.41	5.61
General vocabulary	234	−5.73	1.53	.512	60	−5.82	1.64	57	−5.63	1.55	58	−5.54	1.37	59	−5.92	1.57
Quantitative language	234	8.78	2.39	.056	60	8.80	2.32	57	9.07	2.51	58	9.19	2.55	59	8.08	2.06
Numeracy (full)	234	12.86	6.11	.154	60	12.43	6.14	57	13.96	5.86	58	13.50	6.24	59	11.61	6.09
Numeracy (adjusted)	234	11.36	5.50	.221	60	10.95	5.64	57	12.35	5.32	58	11.81	5.53	59	10.39	5.43
Posttest																
Quantitative language	218	10.39	2.26	.160	59	10.49	2.37	49	10.37	1.88	55	10.84	2.28	55	9.87	2.36
Numeracy (full)	218	16.01	6.28	.204	59	14.81	6.60	49	16.39	5.82	55	17.25	6.15	55	15.73	6.34
Numeracy (adjusted)	218	14.01	5.73	.255	59	12.90	6.03	49	14.31	5.29	55	15.02	5.58	55	13.93	5.85

numeracy items, all analyses were rerun using the PENS measure without the three items that involved questions using more/fewer. Results are largely the same as the primary results and are presented below by hypothesis.

Hypothesis 1 (H1)

There was no significant advantage for children in the quantitative language only condition (compared to active control) for either the quantitative mathematical language outcome (H1a: estimate = 0.33, $p = .139$ [one-tailed], Hedges' $g = 0.15$) or for the numeracy outcome (H1b: estimate = −1.39, $p = .983$ [one-tailed], Hedges' $g = -0.24$).

Hypothesis 2 (H2)

We found that there was no difference between the numeracy only condition and active control condition on numeracy skills (H2a: estimate = −0.74, $p = .859$ [one-tailed], Hedges' $g = -0.13$). The numeracy only condition did not impact quantitative mathematical language (H2b: estimate = −0.05, $p = .856$ [two-tailed], Hedges' $g = -0.02$).

Hypothesis 3 (H3)

We find that those in the combined condition did outperform those in the numeracy only condition on math language (H3a: estimate =

Table 6
Parameter Estimates for Intervention Contrasts Using Dummy-Coded Condition Variables

Predictor	Quantitative math language			Numeracy		
	Estimate	*SE*	*p*	Estimate	*SE*	*p*
Age	0.38	0.22	.086	1.06	0.53	.022
Gender (1 = *female*)	0.23	0.23	.308	−0.51	0.54	.199
Parent education	0.06	0.07	.360	0.33	0.16	.021
Attendance	0.07	0.02	.005	−0.02	0.06	.375
Numeracy skills pretest	0.10	0.03	<.001	0.53	0.06	<.0001
Quantitative language skills pretest	0.24	0.07	<.001	0.37	0.16	.009
Cognitive flexibility	0.07	0.03	.008	0.09	0.06	.069
General vocabulary	0.13	0.09	.126	0.27	0.21	.115
Condition (control as reference)						
Intercept	3.45	1.44	.017	0.99	3.63	.427
Math language condition versus control	0.32	0.30	.148[a]	−1.35	0.72	.970[a]
Numeracy condition versus control	−0.06	0.31	.850	−0.62	0.76	.793[a]
combined condition versus control	0.47	0.31	.126	0.42	0.74	.571
Condition (combined as reference)						
Intercept	3.92	1.44	.007	1.41	3.63	.427
Math language condition versus combined	−0.16	0.29	.598	−1.77	0.71	.019[a,b]
Numeracy condition versus combined	−0.53	0.31	.045[a]	−1.04	0.75	.128[a,b]
Condition (quantitative language as reference)						
Intercept	3.76	1.45	.010	−0.36	3.65	.921
Numeracy versus quantitative language	−0.37	0.30	.218	0.73	0.74	.321

Note. Results are reported across 25 imputations.
[a] A priori directional hypothesis that was tested using a one-tailed test of significance. All other *p*-values were using two-tailed tests because we did not have a priori directional hypotheses. [b] *p* values were adjusted to account for clustering.

−0.54, $p = .042$ [one-tailed], Hedges' $g = -0.24$). However, using the conservative Bonferroni correction, this was no longer statistically significant. We find that the combined condition did not afford an advantage when compared to the numeracy only condition on children's numeracy skills (H3b: estimate = −0.94, $p = .083$ [one-tailed], Hedges' $g = -0.16$). We did, however, see that the combined condition outperformed the quantitative language only condition on numeracy skills (H3c; estimate = −1.58, $p = .048$ [one-tailed], Hedges' $g = -0.28$) which was no longer significant after the use of the conservative Bonferroni correction.

Exploratory Post Hoc Analyses

Given that the effects of the combined condition varied in statistical significance across the primary analyses and robustness check, frequentist analyses do not allow us to support the null hypothesis (that there is no effect), and the effect sizes across these contrasts were consistently positive favoring the combined condition, we calculated Bayes factors (Faulkenberry, 2022) for H3a, H3b, and H3c. The Bayes factors for H3a (effects of combined vs. numeracy conditions on quantitative language skills) provided anecdotal support for the hypothesis ($BF_{10} = 1.39$ [$BF_{10} = 1.46$ using the robustness check contrasts]). H3b (effects of combined vs. numeracy conditions on numeracy skills) indicated anecdotal support for the null hypothesis ($BF_{01} = 1.18$ for both the primary and robustness check contrasts). Finally, Bayes factors indicated moderate evidence supporting H3c (effects of combined vs. quantitative language only conditions on quantitative language skills; $BF_{10} = 6.14$ [$BF_{10} = 5.36$ for the robustness check contrasts]).

Discussion

Understanding how quantitative language and numeracy instruction affect children's mathematical learning is a critical step in advancing instructional opportunities for young children during the preschool period. In the current study, we contrasted four conditions—quantitative language only, numeracy only, combined [quantitative language + numeracy], and an active control—to examine the unique and combined effects of quantitative language and numeracy instruction on children's quantitative mathematical language and numeracy skills. Unfortunately, the effects of the interventions were not as clear, nor as robust, as had been expected based on previous work. Namely, as has been found multiple times in prior research (e.g., Purpura et al., 2021; Purpura, Napoli, et al., 2017) unique effects of the quantitative language only condition on quantitative language skills and numeracy outcomes were not found. Similarly, no effects of the numeracy only intervention were found on either quantitative language or numeracy outcomes. In contrast, the effects of the combined condition were mixed and not as robust as expected.

Unique Effects

Quantitative Language Outcomes

In contrast to prior work, the quantitative language only condition did not statistically significantly outperform the active control condition on quantitative language skills. Although the effect size was positive (Hedges' $g = 0.14$) on quantitative language skills (favoring the quantitative language only condition), it was substantially smaller than the expected effect size from prior work (e.g., Hedges' $g = 0.46$; see Table 1). It is unclear why these results did not align with prior evidence, particularly given that prior evidence on quantitative language only interventions (Purpura et al., 2021; Purpura, Day, et al., 2017; Purpura, Napoli, et al., 2017) showed robust effects on quantitative language skills. However, there were a few notable differences in this intervention study than in the prior studies.

First, in contrast to Purpura et al. (2021) which was implemented one-on-one by parents, this study was conducted in classrooms in small groups of three to four children. The added number of children during the reading may have made it more challenging for children to benefit from the intervention as they may not have had the opportunity to respond to each of the dialogic reading questions that were asked. In fact, if all children in a group of four children were posed an equal number of questions, each child would have the opportunity to respond to only a quarter of the total questions; which, over the course of 24 potential sessions, would be an opportunity to respond to only 102 questions per child in groups of four versus 408 questions when delivered one-on-one. As group sizes in the current study were not randomly determined (they were based on the number of children assigned to each condition within classrooms), we cannot adequately address this question with these data. Future work should more systematically examine group size effects as has been done for other small group interventions (Clarke et al., 2020, 2023).

Second, the effects on quantitative mathematical language may have been dampened due to the relative increase in recognition of the importance of mathematical language over the last several years and potentially the increased attention to it in typical classrooms—which would subsequently raise the baseline of instruction for most students. For example, the recent What Works Clearinghouse report on *Preparing Young Children for School* (Burchinal et al., 2022) includes an emphasis on mathematical language as one of its key recommendations. Similarly, the state in which the intervention was conducted currently includes preschool mathematics standards on knowledge of quantitative mathematical language terms (Indiana Department of Education, 2015). It is possible that the increase in attention to mathematical language more generally by educators may have dampened the effects of the quantitative language only condition by establishing a higher baseline of general mathematical language instruction in classrooms. Further work investigating teachers' systematic use of mathematical language in classrooms is warranted.

Numeracy Outcomes. Unexpectedly, neither the quantitative language only condition nor the numeracy only condition had a positive impact on children's numeracy outcomes. Although we expected positive effects on numeracy outcomes for both conditions, there have been a range of effect sizes on numeracy outcomes from prior research—from nearly one standard deviation below zero (Mix et al., 2012) to around two standard deviations above zero (Mix et al., 2012; Young-Loveridge, 2004). Although these findings are disappointing, they do fall into the range of prior studies. The lack of positive effects could be due to a number of potential factors including breadth of focus of the intervention, the potential need for inclusion of both mathematical language and numeracy instruction for children to acquire numeracy skills, and the increase in attention to mathematical language and numeracy instruction in general classrooms.

First, in much of the prior research focusing on improving numeracy skills through picture book interventions, there has been a very narrow focus on individual numeracy constructs such as cardinality during the intervention as well as the assessments (e.g., Gibson et al., 2020; Mix et al., 2012; O'Rear & McNeil, 2019). The current intervention focused more generally on counting, relations, and operations to better represent the breadth of numeracy skills (Milburn et al., 2019; Purpura & Lonigan, 2013). However, this broader focus (and related measurement) may have diminished the observed effects either because the instructional emphasis was too broad, or because the measurement of impacts was done at a broader level and not at a specific-enough level (i.e., assessment was inclusive of a wide range of numeracy skills when impacts may have been at a more specific level such as just on cardinality). For example, there are few items on the PENS that focus on cardinality, whereas a lot of the text and dialogic reading questions in the numeracy condition are directly aligned with knowledge of exact quantities. Future work should more directly evaluate the intervention effects on individual numeracy constructs.

Second, and building off the prior point, broader numeracy instruction may require quantitative language instruction to be effective. Notably, the numeracy only instruction provided through the books intentionally did not contain any quantitative language. Whereas some prior effective numeracy intervention studies may have naturally included quantitative language terms in instruction depending on who the interventionists were (e.g., parents or teachers vs. trained interventionists)—even if not explicit about it (e.g., prior interventions may have focused on cardinality, but naturally contrasted various quantities using the words more/fewer). Importantly, in discussions with the research team, it was noted that the language used in the numeracy only condition could have been more natural and easier to discuss if there was the embedded use of quantitative language terms. Explicitly removing such terminology and actively working to avoid building on children's natural use of it, may have prevented the normative effectiveness of numeracy instruction.

Third, and similar to recent changes with mathematical language in instructional guidelines, greater attention has been placed on numeracy in preschool classrooms over the last several years. It may be that there was a higher baseline of numeracy focus across all classrooms in the current study than would have been seen previously. For example, in prior work (e.g., Purpura, Day, et al., 2017; Purpura, Napoli, et al., 2017), there was nearly no growth in children's numeracy performance from pretest to posttest for the control condition; whereas, in the current intervention, children in the active control condition gained over 3.64 out of 29 total possible points on the PENS from pretest to posttest. Unfortunately, there are few systematic efforts to measure the amount of mathematics instruction generally over time during the preschool years, so it is difficult to evaluate how instruction has changed over time.

Combined Effects

The effect sizes for H3 analyses were all in the direction favoring the combined condition for all key contrasts, but these effects were mixed and mostly not statistically significant. Although these effects were directionally aligned with our hypotheses, they were of lower magnitude than expected and thus, generally did not reach statistical significance. However, the positive effect sizes shown across both quantitative language and numeracy outcomes, suggest that the combined condition may have more potential compared to instruction in either domain individually to impact student learning.

It should be noted that the positive effect sizes of the combined condition were generated using only half the amount of instructional time in each individual area (quantitative language and numeracy) than either of the other instructional conditions. The total amount of book reading instruction was the same, regardless of condition. Thus, although the combined condition only resulted in an effect size improvement of Hedges' $g = 0.15$ compared to the numeracy only condition, it was done with only half as much time explicitly focused on numeracy instruction. Similarly, the effect sizes of 0.23 on quantitative language (contrasting the combined condition and the numeracy only condition) and 0.28 on numeracy (contrasting the combined condition with the quantitative language only condition) were found with instructional time split across instruction in quantitative language and numeracy. As such, these findings are potentially promising and suggest that more investigation is needed to determine how to more effectively combine and support quantitative language and numeracy instruction to benefit children's mathematical development.

The current study took an approach where quantitative language and numeracy instruction were combined into a single instructional condition and interwoven across the full course of the intervention. However, it is unclear if there are other methods of combining instruction such as more explicitly aligning with learning trajectories (Clements & Sarama, 2011) may be a better method of providing instruction. For example, given that quantitative language instruction may provide children with a basis for acquiring numeracy skills by giving them access to the language used in numeracy instruction (Purpura et al., 2018), combined instruction may be better implemented in an ordered fashion with first providing quantitative language instruction and then providing numeracy instruction—rather than providing both at the same time as done here. Alternatively, similar to recent work that shows the benefits of targeting instruction to children's individual needs based on learning trajectories (Clements et al., 2021) it may be best to provide children with mathematical language and numeracy instruction (or a combination) based on their ability level. Future work should examine how best to maximize children's learning using quantitative language and numeracy instruction.

Limitations

Two key limitations to this study must be noted. First, these interventions were conducted during the height of the Omicron Wave of the COVID-19 pandemic which may have impacted both children's entry knowledge and the consistency of the learning environment during the intervention. In terms of entry knowledge, there is growing evidence that the pandemic impacted children's opportunities to learn (Kuhfeld et al., 2022; K. Lewis & Kuhfeld, 2021) including at the earliest ages through reduced effects of classroom instruction and reduced opportunities to participate in early childhood education programs (Yoshikawa et al., 2020). Unfortunately, we did not comprehensively assess that information in family backgrounds, so we cannot know if it may have affected children's readiness to learn the content. Moreover, there were clear impacts of the Omicron Wave during the intervention implementation that may have disrupted children's opportunities to learn in a consistent learning environment specific to the preschool programs with whom we worked. Notably, for a portion of the intervention period, many of our schools and/or classrooms were implementing 2-week full class quarantines if there was a positive case in a classroom. Although we continued the intervention upon the reopening (at the same spot we left off on), this meant that there could be multiple weeks between intervention sessions—potentially interfering with any benefits that may have been seen and built on. Moreover, when schools discontinued whole classroom quarantines and moved to individual child quarantines, this resulted in a number of children that missed individual sessions and had to be made up when they returned. Each of these challenges led to an extended intervention schedule to complete the targeted number of sessions and a more disrupted school environment for participating children. Thus, these findings may not be generalizable to settings beyond the situation in which this study was implemented due to the unique circumstance under which the intervention was implemented, and future work should further evaluate these contrasts under different environmental conditions.

Second, the use of picture books alone may not be maximally effective in building children's language and numeracy skills. Recent evidence has suggested that vocabulary acquisition (and we would surmise this would apply to mathematical language too) can be enhanced through pairing shared book reading with guided play (Dickinson et al., 2019; Toub et al., 2018). Similarly, there is evidence that numeracy instruction is most effective when done with manipulatives (Carbonneau et al., 2013). Had the interventions expanded to incorporate this research evidence, we may have seen more robust effects of the unique instructional conditions. However, we also note that it may have been more challenging and time-consuming to implement such instruction in the classrooms and more difficult to ensure control of instructional inputs that were necessary in this study.

Future Directions

Given the limitations of the current study and the limited significant findings, we highlight several critical next directions for future research. First, there is a need to develop a more robust set of interventions to maximize impacts on the individual domains of mathematical language and numeracy instruction. This may be done through extending instruction beyond just picture books to include guided play activities and manipulatives. Second, building on the current interventions, there is a need to determine how best to deliver the combined intervention to maximize learning. For example, should quantitative language and numeracy instruction be delivered simultaneously as was done in the current intervention or in sequence according to a learning trajectory? Third, are there individual differences—including initial ability levels—that suggest targeted instruction on either quantitative language or numeracy first (or in combination) may be more effective. Essentially, what type of instruction works best for whom and when? Fourth, although not explicitly intervention-related, it is necessary to better understand how quantitative language and numeracy instruction are occurring in classroom settings. This knowledge will allow us to better target instructional support and identify how the baseline of quantitative language instruction may have changed over time. Fifth, given the impacts of COVID-19 on the intervention timeline of this study, conducting a similar follow-up study without that disruption may help to eliminate whether or not those disruptions adversely affected outcomes.

Conclusion

The findings from the current study were mixed and less robust than expected in evaluating the unique and combined effects of quantitative language and numeracy instruction. Although the unique conditions did not exhibit significant impacts on their expected areas, there were some, more modest, effects of the combined condition on children's learning. These findings highlight the need to better understand how quantitative language and numeracy instruction can be better integrated to enhance children's early learning.

References

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. *Developmental Psychology*, *50*(10), 2368–2379. https://doi.org/10.1037/a0037493

Aunola, K., Leskinen, E., Lerkkanen, M. K., & Nurmi, J. E. (2004). Developmental dynamics of math performance from preschool to grade 2. *Journal of Educational Psychology*, *96*(4), 699–713. https://doi.org/10.1037/0022-0663.96.4.699

Barner, D., & Bachrach, A. (2010). Inference and exact numerical representation in early language development. *Cognitive Psychology*, *60*(1), 40–62. https://doi.org/10.1016/j.cogpsych.2009.06.002

Barner, D., Chow, K., & Yang, S.-J. (2009). Finding one's meaning: A test of the relation between quantifiers and integers in language development. *Cognitive Psychology*, *58*(2), 195–219. https://doi.org/10.1016/j.cogpsych.2008.07.001

Barner, D., Libenson, A., Cheung, P., & Takasaki, M. (2009). Cross-linguistic relations between quantifiers and numerals in language acquisition: Evidence from Japanese. *Journal of Experimental Child Psychology*, *103*(4), 421–440. https://doi.org/10.1016/j.jecp.2008.12.001

Boonen, A. J., Kolkman, M. E., & Kroesbergen, E. H. (2011). The relation between teachers' math talk and the acquisition of number sense within kindergarten classrooms. *Journal of School Psychology*, *49*(3), 281–299. https://doi.org/10.1016/j.jsp.2011.03.002

Burchinal, M., Krowka, S., Newman-Gonchar, R., Jayanthi, M., Gersten, R., Wavell, S., Lyskawa, J., Haymond, K., Bierman, K., Gonzalez, J. E., McClelland, M. M., Nelson, K., Pentimonti, J., Purpura, D. J., Sachs, J., Sarama, J., Schlesinger-Devlin, E., Washington, J., & Rosen, E. (2022). *Preparing young children for school (WWC 2022009)*. National Center for Education Evaluation and Regional Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education. https://whatworks.ed.gov/

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, *105*(2), 380–400. https://doi.org/10.1037/a0031084

Chard, D. J., Baker, S. K., Clarke, B., Jungjohann, K., Davis, K., & Smolkowski, K. (2008). Preventing early mathematics difficulties: The feasibility of a rigorous kindergarten mathematics curriculum. *Learning Disability Quarterly*, *31*(1), 11–20. https://doi.org/10.2307/30035522

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

Clarke, B., Doabler, C. T., Turtura, J., Smolkowski, K., Kosty, D. B., Sutherland, M., Kurtz-Nelson, E., Fien, H., & Baker, S. K. (2020). Examining the efficacy of a kindergarten mathematics intervention by group size and initial skill: Implications for practice and policy. *The Elementary School Journal*, *121*(1), 125–153. https://doi.org/10.1086/710041

Clements, D. H., & Sarama, J. (2011). Early childhood mathematics intervention. *Science*, *333*(6045), 968–970. https://doi.org/10.1126/science.1204537

Clements, D. H., Sarama, J., Baroody, A. J., Kutaka, T. S., Chernyavskiy, P., Joswick, C., Cong, M., & Joseph, E. (2021). Comparing the efficacy of early arithmetic instruction based on a learning trajectory and teaching-to-a-target. *Journal of Educational Psychology*, *113*(7), 1323–1337. https://doi.org/10.1037/edu0000633

Dickinson, D. K., Collins, M. F., Nesbitt, K., Toub, T. S., Hassinger-Das, B., Hadley, E. B., & Golinkoff, R. M. (2019). Effects of teacher-delivered book reading and play on vocabulary learning and self-regulation among low-income preschool children. *Journal of Cognition and Development*, *20*(2), 136–164. https://doi.org/10.1080/15248372.2018.1483373

Dong, N., & Maynard, R. (2013). Powerup!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, *6*(1), 24–67. https://doi.org/10.1080/19345747.2012.673143

Faulkenberry, T. J. (2022). *Psychological statistics: The basics*. Routledge.

Gibson, D. J., Gunderson, E. A., & Levine, S. C. (2020). Causal effects of parent number talk on preschoolers' number knowledge. *Child Development*, *91*(6), e1162–e1177. https://doi.org/10.1111/cdev.13423

Green, K. B., Gallagher, P. A., & Hart, L. (2018). Integrating mathematics and children's literature for young children with disabilities. *Journal of Early Intervention*, *40*(1), 3–19. https://doi.org/10.1177/1053815117737339

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology*, *93*(3), 615–626. https://doi.org/10.1037/0022-0663.93.3.615

Harmon, J. M., Hedrick, W. B., & Wood, K. D. (2005). Research on vocabulary instruction in the content areas: Implications for struggling readers. *Reading and Writing Quarterly*, *21*(3), 261–280. https://doi.org/10.1080/10573560590949377

Hassinger-Das, B., Jordan, N. C., & Dyson, N. (2015). Reading stories to learn math: Mathematics vocabulary instruction for children with early numeracy difficulties. *The Elementary School Journal*, *116*(2), 242–264. https://doi.org/10.1086/683986

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics*, *32*(4), 341–370. https://doi.org/10.3102/1076998606298043

Hedges, L. V., & Hedberg, E. C. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), 60–87. https://doi.org/10.3102/0162373707299706

Hooper, S. R., Roberts, J., Sideris, J., Burchinal, M., & Zeisel, S. (2010). Longitudinal predictors of reading and math trajectories through middle school from African American versus Caucasian students across two samples. *Developmental Psychology*, *46*(5), 1018–1029. https://doi.org/10.1037/a0018877

Hornburg, C. B., Schmitt, S. A., & Purpura, D. J. (2018). Relations between preschoolers' mathematical language understanding and specific numeracy skills. *Journal of Experimental Child Psychology*, *176*, 84–100. https://doi.org/10.1016/j.jecp.2018.07.005

Indiana Department of Education Family and Social Services Administration: Office of Early Childhood and Out of School Learning Early Learning Advisory Committee. (2015). *Indiana early learning foundations*. https://www.in.gov/doe/files/foundations-2015-august-12.pdf

Jennings, C. M., Jennings, J. E., Richey, J., & Dixon-Krauss, L. (1992). Increasing interest and achievement in mathematics through children's literature. *Early Childhood Research Quarterly*, *7*(2), 263–276. https://doi.org/10.1016/0885-2006(92)90008-M

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B., & Irwin, C. (2012). Building kindergartners' number sense: A randomized controlled study. *Journal of Educational Psychology*, *104*(3), 647–660. https://doi.org/10.1037/a0029018

Jordan, N. C., & Hanich, L. B. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities*, *33*(6), 567–578. https://doi.org/10.1177/002221940003300605

Jordan, N. C., Huttenlocher, J., & Levine, S. C. (1994). Assessing early arithmetic abilities: Effects of verbal and nonverbal response types on the calculation performance of middle-and low-income children. *Learning and Individual Differences*, *6*(4), 413–432. https://doi.org/10.1016/1041-6080(94)90003-5

Jordan, N. C., & Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Reviews*, *15*(1), 60–68. https://doi.org/10.1002/ddrr.46

Koepke, K. M., & Miller, B. (2013). At the intersection of math and reading disabilities: Introduction to the special issue. *Journal of Learning Disabilities*, *46*(6), 483–489. https://doi.org/10.1177/0022219413498200

Kuhfeld, M., Soland, J., Lewis, K., Ruzek, E., & Johnson, A. (2022). The COVID-19 school year: Learning and recovery across 2020–2021. *AERA Open*, *8*, Article 23328584221099306. https://doi.org/10.1177/23328584221099306

Lansdell, J. M. (1999). Introducing young children to mathematical concepts: Problems with 'new' terminology. *Educational Studies*, *25*(3), 327–333. https://doi.org/10.1080/03055699997837

LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, *81*(6), 1753–1767. https://doi.org/10.1111/j.1467-8624.2010.01508.x

Lewis, C., Hitch, G. J., & Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9-to 10-year-old boys and girls. *Journal of Child Psychology and Psychiatry*, *35*(2), 283–292. https://doi.org/10.1111/j.1469-7610.1994.tb01162.x

Lewis, K., & Kuhfeld, M. (2021). *Learning during COVID-19: An update on student achievement and growth at the start of the 2021–22 school year*. (NWEA Brief).

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, *50*(4), 315–353. https://doi.org/10.1016/j.cogpsych.2004.09.004

McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders task. *Frontiers in Psychology*, *5*, Article 599. https://doi.org/10.3389/fpsyg.2014.00599

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, *22*(1), 114–140. https://doi.org/10.1037/met0000078

Milburn, T. F., Lonigan, C. J., DeFlorio, L., & Klein, A. (2019). Dimensionality of preschoolers' informal mathematical abilities. *Early Childhood Research Quarterly*, *47*, 487–495. https://doi.org/10.1016/j.ecresq.2018.07.006

Mix, K. S., Sandhofer, C. M., Moore, J. A., & Russell, C. (2012). Acquisition of the cardinal word principle: The role of input. *Early Childhood Research Quarterly*, *27*(2), 274–283. https://doi.org/10.1016/j.ecresq.2011.10.003

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through Grade 8 mathematics*.

O'Rear, C. D., & McNeil, N. M. (2019). Improved set-size labeling mediates the effect of a counting intervention on children's understanding of cardinality. *Developmental Science*, *22*(6), Article e12819. https://doi.org/10.1111/desc.12819

Powell, S. R., & Driver, M. K. (2015). The influence of mathematics vocabulary instruction embedded within addition tutoring for first-grade students with mathematics difficulty. *Learning Disability Quarterly*, *38*(4), 221–233. https://doi.org/10.1177/0731948714564574

Pruden, S. M., Levine, S. C., & Huttenlocher, J. (2011). Children's spatial thinking: Does talk about the spatial world matter? *Developmental Science*, *14*(6), 1417–1430. https://doi.org/10.1111/j.1467-7687.2011.01088.x

Purpura, D. J. (2021). *Preschool early numeracy screener*. Pro-Ed.

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology*, *105*(2), 453–464. https://doi.org/10.1037/a0031753

Purpura, D. J., Day, E., Napoli, A. R., & Hart, S. A. (2017). Identifying domain-general and domain-specific predictors of low mathematics performance: A classification and regression tree analysis. *Journal of Numerical Cognition*, *3*(2), 365–399. https://doi.org/10.5964/jnc.v3i2.53

Purpura, D. J., Gonzalez, J., Anaya, L. Y., Usero-Gonzalez, F., & Kung, M. (2018). The importance of mathematical language in early numeracy development: Implications for dual language learners. In D. Luft Baker, D. Basaraba, & C. Richards-Tutor (Eds.), *Second language acquisition: Methods, perspectives and challenges* (pp. 153–170). Nova Science Publishers.

Purpura, D. J., Hume, L. E., Sims, D. M., & Lonigan, C. J. (2011). Early literacy and early numeracy: The value of including early literacy skills in the prediction of numeracy development. *Journal of Experimental Child Psychology*, *110*(4), 647–658. https://doi.org/10.1016/j.jecp.2011.07.004

Purpura, D. J., & Logan, J. A. (2015). The nonlinear relations of the approximate number system and mathematical language to early mathematics development. *Developmental Psychology*, *51*(12), 1717–1724. https://doi.org/10.1037/dev0000055

Purpura, D. J., & Lonigan, C. J. (2013). Informal numeracy skills: The structure and relations among numbering, relations, and arithmetic operations in preschool. *American Educational Research Journal*, *50*(1), 178–209. https://doi.org/10.3102/0002831212465332

Purpura, D. J., Napoli, A. R., & King, Y. (2019). Development of mathematical language in preschool and its role in learning numeracy skills. In D. C. Geary, D. B. Berch, & K. M. Koepke (Eds.), *Cognitive foundations for improving mathematical learning* (pp. 175–193). Academic Press. https://doi.org/10.1016/B978-0-12-815952-1.00007-4

Purpura, D. J., Napoli, A. R., Wehrspann, E. A., & Gold, Z. S. (2017). Causal connections between mathematical language and mathematical knowledge: A dialogic reading intervention. *Journal of Research on Educational Effectiveness*, *10*(1), 116–137. https://doi.org/10.1080/19345747.2016.1204639

Purpura, D. J., & Reid, E. E. (2016). Mathematics and language: Individual and group differences in mathematical language skills in young children. *Early Childhood Research Quarterly*, *36*, 259–268. https://doi.org/10.1016/j.ecresq.2015.12.020

Purpura, D. J., Schmitt, S. A., Napoli, A. R., Dobbs-Oates, J., King, Y. A., Hornburg, C. B., Westerberg, L., Borriello, G. A., Bryant, L., Anaya, L. Y., Kung, M., Litkowski, E., Lin, J., & Rolan, E. (2021). Engaging caregivers and children in picture books: A family-implemented mathematical language intervention. *Journal of Educational Psychology*, *113*(7), 1338–1353. https://doi.org/10.1037/edu0000662

Ramani, G. B., Zippert, E., Schweitzer, S., & Pan, S. (2014). Preschool children's joint block building during a guided play activity. *Journal of Applied Developmental Psychology*, *35*(4), 326–336. https://doi.org/10.1016/j.appdev.2014.05.005

Romano, E., Babchishin, L., Pagani, L. S., & Kohen, D. (2010). School readiness and later achievement: Replication and extension using a nationwide Canadian survey. *Developmental Psychology*, *46*(5), 995–1007. https://doi.org/10.1037/a0018880

Rothbart, M. K., & Posner, M. I. (2005). Genes and experience in the development of executive attention and effortful control. *New Directions for Child and Adolescent Development*, *109*, 101–108. https://doi.org/10.1002/cd.142

Schmitt, S. A., Korucu, I., Napoli, A. R., Bryant, L. M., & Purpura, D. J. (2018). Using block play to enhance preschool children's mathematics and executive functioning: A randomized control trial. *Early Childhood Research Quarterly*, *44*, 181–191. https://doi.org/10.1016/j.ecresq.2018.04.006

Spelke, E. S. (2003). What makes humans smart? Core knowledge and natural language. In D. Gentner, & S. Goldin-Meadow (Eds.), *Language in mind* (pp. 277–311). MIT Press.

Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, *19*(1), 99–120. https://doi.org/10.1016/j.ecresq.2004.01.002

Toll, S. W., & Van Luit, J. E. (2014). The developmental relationship between language and low early numeracy skills throughout kindergarten. *Exceptional Children*, *81*(1), 64–78. https://doi.org/10.1177/0014402914532233

Toub, T. S., Hassinger-Das, B., Nesbitt, K. T., Ilgaz, H., Weisberg, D. S., Hirsh-Pasek, K., Golinkoff, R. M., Nicolopoulou, A., & Dickinson, D. K. (2018). The language of play: Developing preschool vocabulary through play following shared book-reading. *Early Childhood Research Quarterly*, *45*, 1–17. https://doi.org/10.1016/j.ecresq.2018.01.010

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., Havlik, R., Conway, K., Edwards, E., Freund, L., King, J. W., Moy, C., Witt, E., & Gershon, R. C. (2013). Cognition assessment using the NIH toolbox. *Neurology*, *80*(11 Suppl. 3), S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded

What Works Clearinghouse. (2020). *What works clearinghouse standards handbook* (Version 4.1). Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/ncee/wwc/handbooks

What Works Clearinghouse. (2022). *What works clearinghouse procedures and standards handbook* (Version 5.0). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Handbooks

Whitehurst, G. J. (1992). *Dialogic reading: An effective way to read to preschoolers*. Reading Rockets. https://www.readingrockets.org/article/dialogic-reading-effective-way-read-preschooler

Yoshikawa, H., Wuermli, A. J., Britto, P. R., Dreyer, B., Leckman, J. F., Lye, S. J., Ponguta, L. A., Richter, L. M., & Stein, A. (2020). Effects of the global coronavirus desease-2019 pandemic on early childhood development: Short- and long-term risks and mitigating program and policy actions. *The Journal of Pediatrics*, *223*, 188–193. https://doi.org/10.1016/j.jpeds.2020.05.020

Young-Loveridge, J. M. (2004). Effects on early numeracy of a program using number books and games. *Early Childhood Research Quarterly*, *19*(1), 82–98. https://doi.org/10.1016/j.ecresq.2004.01.001

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*(1), 297–301. https://doi.org/10.1038/nprot.2006.46

(*Appendix follows*)

Appendix

Previous Mathematical Language and Numeracy Intervention Studies

Intervention study	Condition	Outcome	Condition 1 (*n*)	Condition 2 (*n*)	Hedges' *g* effect size
Gibson et al. (2020)	Numeracy versus control	Numeracy	37	31	0.17
Green et al. (2018)	Q + N versus control	Numeracy	24	26	0.65
Green et al. (2018)	Q + N versus control	Numeracy	24	26	0.54
Green et al. (2018)	Q + N versus control	Numeracy	24	26	0.47
Green et al. (2018)	Q + N versus control	Numeracy	24	26	0.57
Hassinger-Das et al. (2015)	Q + N versus numeracy	Language	42	42	0.32
Hassinger-Das et al. (2015)	Q + N versus control	Language	42	40	0.17
Hassinger-Das et al. (2015)	Numeracy versus control	Language	42	40	−0.17
Hassinger-Das et al. (2015)	Q + N versus numeracy	Language	42	42	0.41
Hassinger-Das et al. (2015)	Q + N versus control	Language	42	40	0.21
Hassinger-Das et al. (2015)	Numeracy versus control	Language	42	40	−0.21
Hassinger-Das et al. (2015)	Q + N versus numeracy	Language	42	42	0.37
Hassinger-Das et al. (2015)	Q + N versus control	Language	42	40	0.33
Hassinger-Das et al. (2015)	Numeracy versus control	Language	42	40	−0.05
Hassinger-Das et al. (2015)	Q + N versus numeracy	Language	42	42	0.95
Hassinger-Das et al. (2015)	Q + N versus control	Language	42	40	0.81
Hassinger-Das et al. (2015)	Numeracy versus control	Language	42	40	−0.15
Hassinger-Das et al. (2015)	Q + N versus numeracy	Numeracy	42	42	−0.09
Hassinger-Das et al. (2015)	Q + N versus control	Numeracy	42	40	0.40
Hassinger-Das et al. (2015)	Numeracy versus control	Numeracy	42	40	0.47
Hassinger-Das et al. (2015)	Q + N versus numeracy	Numeracy	42	42	0.00
Hassinger-Das et al. (2015)	Q + N versus control	Numeracy	42	40	0.24
Hassinger-Das et al. (2015)	Numeracy versus control	Numeracy	42	40	0.26
Hassinger-Das et al. (2015)	Q + N versus numeracy	Numeracy	42	42	−0.56
Hassinger-Das et al. (2015)	Q + N versus control	Numeracy	42	40	0.09
Hassinger-Das et al. (2015)	Numeracy versus control	Numeracy	42	40	0.69
Jennings et al. (1992)	Q + N versus control	Numeracy	29	32	0.40
Jennings et al. (1992)	Q + N versus control	Numeracy	29	32	0.03
Jennings et al. (1992)	Q + N versus control	Numeracy	29	32	1.41
Mix et al. (2012)	Numeracy versus control	Numeracy	15	15	0.48
Mix et al. (2012)	Numeracy versus control	Numeracy	15	15	0.00
Mix et al. (2012)	Numeracy versus control	Numeracy	15	15	−0.85
Mix et al. (2012)	Numeracy versus control	Numeracy	15	15	−0.07
Mix et al. (2012)	Numeracy versus control	Numeracy	15	15	1.15
Mix et al. (2012)	Numeracy versus control	Numeracy	15	15	2.13
Mix et al. (2012)	Numeracy versus control	Numeracy	15	15	2.13
Mix et al. (2012)	Numeracy versus control	Numeracy	15	15	1.03
Mix et al. (2012)	Numeracy versus control	Numeracy	15	15	0.72
O'Rear and McNeil (2019)	Numeracy versus control	Numeracy	34	37	0.72
O'Rear and McNeil (2019)	Numeracy versus control	Numeracy	35	37	0.55
Powell and Driver (2015)	Q + N versus control	Language	35	28	0.48
Powell and Driver (2015)	Q + N versus numeracy	Language	35	35	−0.16
Powell and Driver (2015)	Numeracy versus control	Language	35	28	0.63
Powell and Driver (2015)	Q + N versus control	Numeracy	35	28	0.19
Powell and Driver (2015)	Q + N versus numeracy	Numeracy	35	35	−0.31
Powell and Driver (2015)	Numeracy versus control	Numeracy	35	28	0.47
Purpura et al. (2021)	Quant versus control	Language	40	44	0.49
Purpura et al. (2021)	Quant versus control	Numeracy	40	44	0.22
Young-Loveridge (2004)	Numeracy versus control	Numeracy	23	83	1.90

Note. Q + N = quantitative language and numeracy condition. Language outcome is mathematical language. Although each of these studies uses different assessment tools, effect sizes from each study were converted to Hedges' *g* by the authors so they were on the same metric.

Received February 2, 2021
Revision received July 18, 2023
Accepted July 20, 2023 ■