Mathematics-writing synthesis: Kindergarten through Grade 12 mathematics-writing outcomes and instructional methods

Tessa L. Arsenault[2] [ORCID] · **Sarah R. Powell[1]** [ORCID] · **Sarah G. King[1]** [ORCID]

Abstract

In the last decade, researchers have focused more on how to provide instructional supports for mathematics writing (Powell et al., 2017). In this synthesis, we examined 22 studies about mathematics writing to determine the overall mathematics-writing and mathematics outcomes of mathematics-writing instruction, the mathematics-writing and mathematics outcomes based on study and instructional features, the methods used within mathematics-writing instruction, and the methods used within studies involving students with mathematics difficulty (MD). We conducted a comprehensive review of peer-reviewed studies, published since 2000, focused on instruction in mathematics writing in Kindergarten through Grade 12. We included studies with and without students with MD. Results demonstrated positive student outcomes for mathematics-writing instruction. Furthermore, instructional methods for mathematics writing frequently aligned with practices used for mathematics instruction and writing instruction. Although we identified positive outcomes related to mathematics-writing instruction, and we would recommend for practitioners to provide instruction in mathematics writing, results should be interpreted with caution. Additional high-quality research on mathematics-writing instruction is needed to verify and extend on the results from this synthesis.

Keywords Mathematics · Mathematics writing · Mathematics difficulty · Review · Synthesis

✉ Tessa L. Arsenault
  tarsenault@utexas.edu

1   The University of Texas at Austin, Austin, USA

2   The University of Texas at El Paso, El Paso, USA

🖄 Springer

Introduction

Literacy in mathematics involves the ability to reason and communicate both orally and in writing (Dunston & Tyminski, 2013). According to the Common Core State Standards, students must build logical statements supported by data to communicate with others (National Governors Association Center for Best Practices & Council of Chief State School Officers 2010). Additionally, the inclusion of mathematics-writing prompts on high-stakes tests to evaluate reasoning and communication in mathematics has become common in recent years. In fact, nearly half of states use mathematics-writing prompts for Grades 4 and 5 on high-stakes mathematics assessments (Powell & Hebert, 2022). Due to the increasing focus on mathematics writing within standards and assessments, researchers have begun to examine how students write in mathematics and how to support students in mathematics writing (Arsenault et al., 2022; Hughes et al., 2020; Powell et al., 2017). Although such research indicates educators report the use of mathematics writing in the classroom holds value to improve mathematics outcomes and writing outcomes (Powell et al., 2021), little research examines the efficacy of mathematics writing instructional practices across studies (Powell et al., 2017).

In this synthesis, we reviewed research studies in which authors provided instruction on mathematics writing for students in Kindergarten through Grade 12. We examined the mathematics writing and mathematics outcomes of mathematics-writing instruction, the instructional components used to teach mathematics writing, and if identified instructional components differed for students with and without mathematics difficulty (MD). In the introduction, we define the types of mathematics writing (i.e., exploratory, informative, argumentative, and mathematically creative; Casa et al., 2016). Next, we review frequent methods for supporting students in mathematics writing. Then, we present the challenges of mathematics writing for all students and for students with MD. Finally, we present the purpose and research questions of the synthesis.

Defining mathematics writing

Educators perceive mathematics writing as a valuable component of mathematics. In fact, over half of educators reported including mathematics writing in the classroom at least once a week (Powell et al., 2021). The inclusion of mathematics writing in the classroom targets the development and assessment of student understanding of mathematical concepts (Casa et al., 2016; Powell et al., 2021). For this synthesis, we defined *mathematics writing* as the learning and assessment of student mathematical understanding in the four mathematics-writing categories outlined by the Elementary Mathematical Writing Task Force: exploratory, informative, argumentative, and mathematically creative (Casa et al., 2016).

Exploratory mathematics writing

Within exploratory mathematics writing, students act as their own audience to support their understanding of mathematical concepts (Casa et al., 2016). A prompt for exploratory writing may require a student to write about their difficulties while working through a mathematics problem and how they overcame the mathematics difficulty (Tan & Garces-Bacsal, 2016). The use of exploratory writing can support students to develop mathematical ideas as well as lead to engaging students in other types of mathematics writing (Casa et al., 2016).

Informative mathematics writing

For informative mathematics writing, an educator, or another student, acts as the audience as students write explanations about a mathematical concept (Casa et al., 2016). For example, an informative prompt may provide the work of a pseudo student with a prompt for the student to identify the mistakes and explain how they would solve the problem correctly (Namkung et al., 2019). Powell et al. (2017) reported three-fourths of intervention studies focused on informative mathematics writing. When surveyed, 60% of Kindergarten through Grade 12 educators reported that they required students to explain their work through informative mathematics writing at least once a week (Powell et al., 2021).

Argumentative mathematics writing

In argumentative writing, an educator, or another student, acts as the audience as the student constructs an argument or critiques the reasoning of others (Casa et al., 2016). In argumentative writing prompts, students write a claim based on solving a problem and defend the truth of their claim (Kosko & Zimmerman, 2019). The use of argumentative writing becomes especially important as students write mathematics proofs in geometry (Fuentes, 2011). Powell et al. (2021) reported that 51% of educators require students to write arguments in mathematics.

Mathematically creative writing

In mathematically creative writing, students write for a wider audience to document original mathematical ideas or create problems. For mathematically creative writing prompts, students may create their own word problems or mathematical stories (Casa et al., 2016; Levenberg, 2014). Across grade levels, only 36% of educators reported the use of mathematically creative writing in the classroom (Powell et al., 2021).

Supporting mathematics writing

Previous research indicates the potential benefits of instruction and practices opportunities for students engaging in mathematics writing. Educators may provide mathematics-writing instruction as a primary focus of instruction or as an integrated approach to supporting mathematics learning (Baxter et al., 2005; Swanson et al., 2019). When integrating mathematics writing into mathematics instruction, educators frequently have included journal writing, letter writing, explaining problem solving, writing mathematics-vocabulary definitions, and writing word problems (Powell et al., 2017, 2021). Many of which may not require a tremendous amount of educator support for students to engage in them successfully. When providing targeted mathematics-writing instruction, educators reported including explanations of mathematics writing, modeling, opportunities for practice in mathematics writing, and feedback about mathematics writing (Powell et al., 2021).

Although many U.S. students use mathematics writing on high-stakes assessments, and educators include mathematics writing in instruction (Powell & Hebert, 2022; Powell et al., 2021), little consensus exists on how to effectively support mathematics writing (Powell et al., 2017). Across studies with mathematics-writing instruction, high variability exists in reporting student outcomes, such as only providing data about mathematics outcomes or mathematics-writing outcomes or by only providing descriptions of mathematics-writing features without formal scoring (Powell et al., 2017). Within the evidence base, variability exists across outcome measures that target mathematics-writing instruction, and inconsistencies across measures have made it challenging to draw conclusions about the efficacy of mathematics-writing instruction. With this synthesis, we aimed to provide a comprehensive overview of mathematics-writing instructional practices to inform teaching and research related to mathematics writing.

To date, we identified one synthesis about mathematics writing (Powell et al., 2017). The authors identified 29 studies in which mathematics writing was featured in instruction, as academic assessments, or surveys. With their synthesis, they were most interested in surveying the empirical research base of mathematics writing to determine how many mathematics-writing studies have been published. Of all 29 studies, 17 provided information about mathematics-writing instruction. Powell et al. (2017) coded each study for the mathematics content, mathematics-writing type (e.g., informative, argumentative), implementer, type of assessment, and results. The authors noted that most studies used informative mathematics writing with journal writing as the most popular method for writing in mathematics. In only seven studies did educators ask students to engage in organized classroom writing (i.e., instruction) with only a handful of studies collecting data about the efficacy of mathematics-writing instruction. Powell et al. (2017) did not evaluate the quality of the studies, literature focused on students with MD, or grey literature, such as dissertations, which would be important given the exploratory nature of many studies of mathematics writing. Furthermore, the authors suggested they had difficulty drawing conclusions about the

efficacy of mathematics writing because of the sparse amount of data available and suggested the necessity of future research in all areas of mathematics writing.

Challenges of mathematics writing

To successfully communicate and reason through mathematics writing, students must use general writing skills, computation, mathematics language, and (on occasion) visual representations within their writing (Arsenault et al., 2022; Casa et al., 2016; Hebert & Powell, 2016; Hughes et al., 2020; Powell & Hebert, 2016). However, mathematics writing continues to place a strain on students with and without MD. In the following paragraphs we outline each of the prerequisite skill areas and describe associated challenges.

Within general writing skills, planning and organization are especially challenging for students in mathematics writing (Correnti et al., 2013; Hebert & Powell, 2016). In informative mathematics writing, students must organize their writing with an introduction and conclusion in addition to the problem-solving procedures (Hughes et al., 2020). For argumentative mathematics writing, students must also include a rebuttal (Hughes et al., 2020). Although including an introduction, conclusion, and rebuttal, if needed, supports accurate responses, often, students only include the problem-solving features (Hebert & Powell, 2016). Students may include an introduction sentence, but they rarely write a concluding sentence (Correnti et al., 2013; Hebert & Powell, 2016). Therefore, the development of general writing skills can help effectively convey mathematical ideas.

A second challenge for mathematics writing may include doing computations within a problem. Mathematics-writing prompts frequently include a prompt for students to review computations completed by a pseudo student or prompt students to complete computation problems themselves (Hebert & Powell, 2016; Hughes & Lee, 2019). Yet, students frequently make computation errors or do not attempt computation problems to respond to mathematics-writing prompts (Arsenault et al., 2022; Hebert & Powell, 2016). When responding to informative mathematics-writing prompts with a pseudo student, only 38.9% of students attempted to check the student's work by doing the computation themselves (Arsenault et al., 2022). Additionally, only one-third of students wrote equations in their mathematics-writing responses. When students included equations in their mathematics-writing responses, only 32% correctly set up their equations (Hebert & Powell, 2016). Mathematics writing cannot be separated from computational understanding. The connection likely relates to the necessity for students to be able to complete the computations with accuracy to then express mathematical understanding in mathematics writing. Therefore, when supporting mathematics writing, student computational skills must be considered.

Another hurdle for mathematics writing may be related to the use of mathematics vocabulary. To perform well on mathematics-writing prompts, students must include specific and clear mathematics vocabulary (Stonewater, 2002). Precise and clear mathematics vocabulary includes technical (e.g., integer, quadrilateral), subtechnical (i.e., degrees), general (i.e., difference, more), and

symbolic vocabulary (i.e., +, 5; Monroe & Panchyshyn, 1995). The inclusion, or exclusion, of specific mathematics vocabulary impacts quality of mathematics-writing responses (Hebert & Powell, 2016; Hughes et al., 2020; Stonewater, 2002).

Another challenge may be the use of visual representations within mathematics writing. The inclusion of visual representations supports clear mathematics writing by going beyond words to communicate mathematically and can indicate a high-level of understanding of a problem (Casa et al., 2016; Utami et al., 2019). The inclusion of visual representations correlates with the production of correct mathematics-writing responses (Hughes et al., 2020). Yet, students frequently do not include visual representations in their writing, especially if the prompt did not include a visual representation (Arsenault et al., 2022; Hebert & Powell, 2016).

Students with mathematics difficulty (MD)

We also focus on mathematics writing for students with MD. Students with MD perform below grade-level performance or below the average range in mathematics based on researcher assessments or school diagnosis (Nelson & Powell, 2018), including both students with Individualized Education Programs with goals in mathematics as well as students who perform below a cut-off score assigned by researchers. Cut-off scores frequently range between the 10th to 45th percentile (Geary et al., 2012; Hecht & Vagi, 2010). Students with MD typically perform below their same aged peers across early numeracy, computation, rational number, and word-problem skills (Arsenault & Powell, 2022; Nelson & Powell, 2018).

For students with MD, mathematics writing can be an especially challenging task because it requires students to access both general writing and mathematical prerequisite skills, something that can challenge all students (Arsenault et al., 2022; Hebert & Powell, 2016; Hughes et al., 2020). On mathematics-writing tasks, students with MD typically perform lower than peers on mathematics-writing prompts (Arsenault et al., 2022). Additionally, students with MD write fewer words, numbers, and symbols than their typically achieving peers when responding to mathematics-writing prompts (Arsenault et al., 2022; Hebert & Powell, 2016).

When writing in mathematics, students with MD demonstrate difficulty with general writing skills as well as mathematics specific skills (Hughes et al., 2020). Hughes et al. (2020) reported that when responding to a mathematics-writing prompt, students with MD experienced difficulty with general writing skills as they tried to explain computational procedures and engage in reasoning. Students with MD also had trouble with completing computation problems within mathematics-writing prompts (Hughes et al., 2020). For example, when responding to an informative mathematics-writing prompt with a fraction word problem, only 28 out of 51 students wrote the correct answer to the fraction problem (Hughes et al., 2020). Frequently, students with MD will not even attempt to complete needed computations (Arsenault et al., 2022).

Purpose and research questions

Educators perceive mathematics writing as valuable for supporting mathematics understanding (Powell et al., 2017, 2021). Research also indicates that mathematics writing can be included in the classroom in a variety of formats, but only one past synthesis examines mathematics-writing practices for supporting students' mathematics-writing performance (Powell et al., 2017). Considering the growing importance of mathematics writing and the challenges students frequently experience with mathematics writing, especially for students with MD, it is important that an updated synthesis examines student outcomes after participating in mathematics-writing instruction, particularly for students with MD. It is also necessary to investigate the quality of mathematics-writing studies and all data related to mathematics-writing and general-mathematics outcomes of mathematics-writing instruction to understand the richness of literature related to mathematics-writing instruction.

In this synthesis, we reviewed the Kindergarten to Grade 12 studies focused on either (a) mathematics-writing instruction or (b) mathematics instruction with mathematics writing as a component of instruction. We investigated the following research questions:

1. How do students who participate in mathematics-writing instruction perform on measures of mathematics writing and general mathematics?
2. Do mathematics-writing and general-mathematics outcomes differ across study and instructional features of mathematics-writing instruction (i.e., type of mathematics writing, instructional focus, participant type)?
3. What methods are used to practice mathematics writing during mathematics-writing instruction?
4. For the studies including students with MD, what methods are used to practice mathematics writing during mathematics-writing instruction?

Methods

Search procedures

We conducted a comprehensive search of the literature dated January 2000 to June 2022 to select studies measuring mathematics instruction with mathematics writing as the main instructional goal or as a component of the instruction. We selected 2000 as the start date because of the National Council of Teachers of Mathematics standards released in 2000. This marked a transition point for instructional planning due to the increased focus on rigorous standards in mathematics for Grades K to 12 (National Council of Teachers of Mathematics, 2014). We searched three databases: PyscINFO, Education Source, and ERIC. For the search, we used one line of search terms: "written math*" OR "math* writing" OR ("word problem*" AND math* AND writ*) OR ("open response*" AND math* AND assessment) OR ("open ended" AND math* AND assessment) OR (math* AND (writ* n2 assess*)) OR (calculus AND writing) OR (geometry AND writing) OR (algebra AND writing)
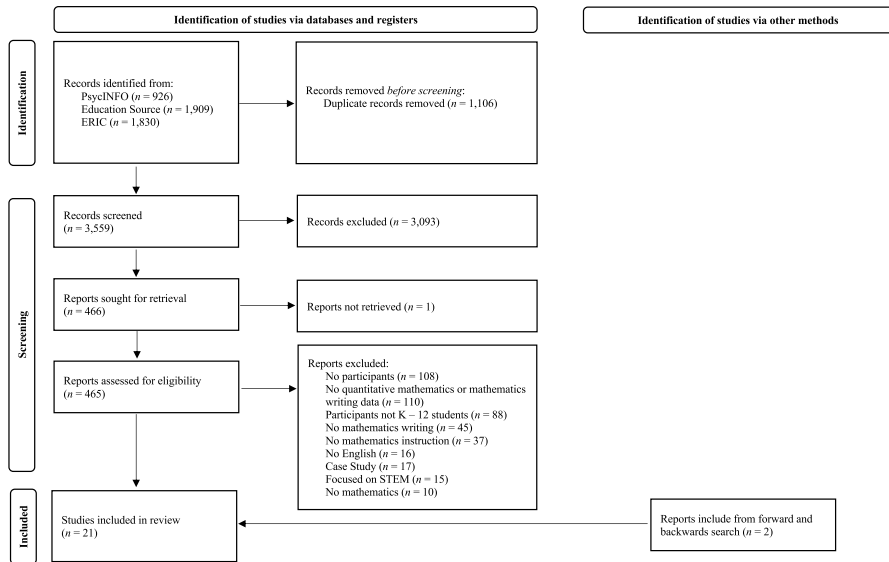
Identification of studies via databases and registers

Identification of studies via other methods

Identification

Records identified from:
PsycINFO ($n = 926$)
Education Source ($n = 1,909$)
ERIC ($n = 1,830$)

Records removed *before screening*:
Duplicate records removed ($n = 1,106$)

Screening

Records screened
($n = 3,559$)

Records excluded ($n = 3,093$)

Reports sought for retrieval
($n = 466$)

Reports not retrieved ($n = 1$)

Reports assessed for eligibility
($n = 465$)

Reports excluded:
No participants ($n = 108$)
No quantitative mathematics or mathematics writing data ($n = 110$)
Participants not K – 12 students ($n = 88$)
No mathematics writing ($n = 45$)
No mathematics instruction ($n = 37$)
No English ($n = 16$)
Case Study ($n = 17$)
Focused on STEM ($n = 15$)
No mathematics ($n = 10$)

Included

Studies included in review
($n = 21$)

Reports include from forward and backwards search ($n = 2$)

Fig. 1 PRISMA diagram

OR TI(math* AND writ*) OR AB(math* n2 writ*) OR SU(mathematics AND writing) OR SU(mathematics AND "written communication"). Figure 1, the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) flowchart (Page et al., 2021), shows the results of the search and the screening of the artifacts. The initial search resulted in 4665 artifacts, decreasing to 3559 artifacts after deduplication. Next, we reviewed titles and abstracts, excluding 3093 artifacts, leaving 466 for full text screening. During full text screening, we excluded 446 artifacts, resulting in 19 articles that met the inclusion criteria. Next, we conducted a forward and backwards search on the included articles. We included one article from the forward search and one article from the backwards search. We also conducted a forward and backwards search on the two additional articles, but we identified no further articles for inclusion. After the forward and backwards search, 21 articles met inclusion criteria for the synthesis.

Inclusion and exclusion criteria

To be included in this synthesis, studies had to meet the following six criteria: (a) The study was published in English. (b) The study was published in or after the year 2000 through 2022. (c) The study design was a randomized control trial, quasi-experimental, regression discontinuity, or single case. For quasi-experimental design studies, treatment and control groups were required, pretests were completed, and a method of pretest equivalence was reported. Table 3 reports the pretest equivalence for each study with a quasi-experimental design. We included these four designs due to previous reports of the limited number of studies including mathematics-writing

instruction (Powell et al., 2017). By including a wide range of study designs, we gained a broad representation of the effects and methods in studies with mathematics-writing instruction. (d) The study included mathematics writing instruction. We defined mathematics writing in four categories: exploratory, informative, argumentative, and mathematically creative (Casa et al., 2016). We defined mathematics writing instruction to include educator explaining, educator modeling, educator demonstrating, and/or student practice in mathematics writing (Powell et al., 2021). (e) The authors include a measure of mathematics writing or mathematics for all students who participated in the instruction (except for attrition). Measures of mathematics writing data included quantitative mathematics writing assessments where students wrote mathematically to respond to prompts or questions or mathematics writing artifacts coded for features included or with a rubric (Powell et al., 2017). Measures of mathematics included quantitative mathematics assessments or mathematics artifacts coded with a rubric. (f) The study was conducted with school-aged students from Kindergarten to Grade 12. (g) The study was published in a peer-reviewed journal.

We excluded articles when they met the following exclusion criteria: (a) No participants in the article. For example, when the study referenced a vignette without actual student participants (e.g., Fello & Paquette, 2009). (b) The article included no quantitative mathematics writing or other mathematics assessment data. For instance, Alvi and Nausheen (2019) included interview data, student engagement data, and qualitative data describing student problem-solving episodes, but we excluded the study due to no quantitative mathematics writing or other mathematics assessment data. (c) The participants in the study were outside the Kindergarten to Grade 12 range, such as when student data was not disaggregated from educator data (e.g., Kosko & Norton, 2012). (d) The study did not include mathematics writing, according to our operational definition. For example, if study defined mathematics writing as handwritten algorithms or writing a word problem rather than exploratory, informative, argumentative, or creative writing in mathematics, we excluded the study (e.g., Broto & Greer, 2014). (e) The study did not include mathematics-writing instruction. For instance, Kosko and Zimmerman (2019) only included a mathematics-writing assessment without instruction. (f) The study was a case-study design with only a treatment group and no control group or did not include pretest equivalence between groups (e.g., Baxter et al., 2005). (g) The study was written in a language other than English (e.g., Çontay & Duatepe-Paksu, 2018). (h) The instruction focused on science, technology, or engineering with mathematics. For example, Casler-Failing (2018) was excluded because of the focus on robotics with mathematics. (i) The study did not include mathematics (e.g., van Drie et al., 2005).

Coding procedures

We coded the 21 articles identified for the synthesis for study demographics, mathematics-writing category, mathematics category, intervention characteristics, student outcomes, and study quality. For study demographics, we coded for study design (i.e., randomized control trial, quasi-experimental, regression discontinuity, or

single case), study location (i.e., country), student grade level and age, student gender (i.e., female or male), student race or ethnicity, number of students in treatment and control groups, sample MD status, and English language status. For sample MD status, we coded the percent of students with MD included in the sample. If the authors reported the percent of students with MD, we coded the sample as "Yes" and included the exact percent of students with MD included. We coded the sample as "No" for MD if the authors did not report the percentage of students with MD. For example, if the authors reported that the sample included a class of students but did not report how many of the students were students with MD.

For mathematics categories, we documented all types of mathematics content areas covered in the study. We included: early numeracy, algebra, fractions, geometry, life skills, measurement, operations, word-problem solving, pre-algebra and algebra, and calculus. We also included an option for "other" to fill in alternative mathematics categories not listed.

For mathematics-writing categories, we classified the mathematics writing completed in each study as exploratory, informative, argumentative, or mathematically creative (Casa et al., 2016). Samples of exploratory mathematics writing included when students wrote to make sense of their own thoughts about mathematics. We coded studies to include mathematics informative writing if the student provided information or explained mathematics concepts. Writing qualified as argumentative mathematics writing if the students used writing to construct mathematical arguments and critique the reasoning of others. Last, mathematically-creative writing included writing samples when students wrote creatively to communicate original ideas, problems with written responses, or solutions (Casa et al., 2016).

With the mathematics-writing instruction, we focused on instructional features and mathematics-writing instructional methods. The instructional features included total instruction time in minutes (calculated through provided information on number of sessions, duration of sessions, frequency of instruction), group size (i.e., whole class, small groups between 2 and 8 students, small groups with size of group not reported, individual), and implementer (educator, researcher, preservice educator, peer tutor, other). We also coded for the purpose of instruction: a goal of increasing mathematics-writing performance or a goal to improve mathematics performance. We defined studies as focused on mathematics writing if the study purpose included a statement focused on mathematics-writing performance, it was the only instructional component, or if there was modeling with practice in mathematics writing. We defined studies as focused on mathematics content with a mathematics-writing component if only one component of a multi-component mathematics intervention included mathematics writing. Next, we coded for the mathematics-writing instructional methods: self-regulated strategy development (SRSD; Graham et al., 2005), attack strategy, paraphrasing, journal writing, letter writing, defining vocabulary, note taking, responding to mathematics-writing prompts, explaining after solving a mathematics problem, argumentative writing, modeling of mathematics writing, or other mathematics-writing practice (see Table 1 for operational definitions).

Next, we coded for the measure information and student outcomes. We recorded the type of measures (i.e., mathematics test, mathematics-writing test, mathematics-vocabulary test, reading test on mathematics) and the timing of the measure (i.e.,

Table 1 Operational definitions of instructional methods

Instructional methods	Operational definition
Argumentative writing	Student solving a problem and justifying their mathematics problem solving
Attack strategy	A step-by-step strategy to work through the mathematics and/or mathematics writing process
Compared writing samples	Students examined the difference between two mathematics writing samples
Created problems	Mathematical word problems written by students
Defining vocabulary	Writing or verbally defining vocabulary individually or as a class
Discussion	Educator to students or students to students discussing mathematical writing topic
Drawing	Students creating visual representations
Feedback	Educator providing feedback to students on mathematics or mathematics-writing work
Formal mathematics-writing prompt	A writing prompt focused on mathematics content
Graphic organizer	A visual display to organize mathematical or written ideas to plan for mathematics and/or writing
Informative writing	Student solving a problem and then explaining how they solved the problem
Journal writing	Any mathematics writing (i.e., exploratory, informative, argumentative, creative) in a paper or online journal
Letter writing	Students writing letters about mathematical topics
Modeling	Educator demonstrating mathematics writing for students
Note taking	Educator writing for students to see and students copying the information or the educator speaking and students writing the information
Paraphrasing	Summarizing information from a problem or other text
Questioning	Educator or student asking questions
Responding to other's work	Students responding to work and questions of other students
Revising	Editing writing to improve clarity and precision
Self-evaluation	Students evaluated their own writing based on a mathematical writing rubric
SRSD	Explicit writing instruction with a focus on self-regulation procedures including goal setting, self-monitoring, self-instructions, and self-reinforcement (Graham et al., 2005)
Story writing	Students writing stories on mathematical topics
Think aloud	Student talking aloud to work through a mathematical writing topic
Writing in response to discussion	Writing after a discussion on a mathematical topic
Writing practices	Writing on topics not related to mathematics
Word wall	Terms and representations of terms posted on a wall for students to reference while participating in mathematical writing

SRSD, self-regulated strategy development

completed during instruction, pretest/posttest). We also coded the type of comparison (i.e., pretest/posttest, treatment/control, baseline/treatment). We also documented the mathematics content area (i.e., early numeracy, algebra, fractions, geometry, life skills, measurement, operations, word-problem solving, pre-algebra/algebra, calculus, other). For the mathematics-writing measures, we recorded mathematics-writing type (i.e., exploratory, informative, argumentative, mathematically creative, or other; Casa et al., 2016). For mathematics-vocabulary and reading measures on mathematics, we coded for the content description. Finally, for all measure types, we coded the assessment name and if applicable the subcategory names.

After documenting the measure information, we coded for the student outcomes on all mathematics and mathematics-writing measures. For studies to include pretest and posttest outcomes (i.e., randomized control trials and quasi-experimental design), we coded the means with provided units, standard deviations, and group sizes at pretest and posttests for all groups. We also recorded significance testing and effect sizes (i.e., Cohen's *d*, Eta Squared, and Hedges' *g*) when reported. For single-case design studies, we recorded progress monitoring data during baseline and instruction. We also included any pretest and posttest means and standard deviations, mean and range of non-overlap for all pairs (NAP), effect sizes (i.e., Tau-U), confidence intervals, and significance testing when reported. Finally, we coded for study quality. We used the Cook et al. (2015) quality indicator checklist; measuring study quality for both group design studies and single-case design studies.

Reliability coding

The first author coded all studies. The third author double-coded 24% of the studies in the synthesis. The first author trained the third author on the coding sheet and manual in a 1-h training. During the training, the first and third author coded a practice article. Then, the third author independently double-coded five of the articles in the synthesis. On these studies, we obtained an overall inter-rater reliability of 80%. The first and third author then met to discuss any disagreements to reach consensus. We reached 100% inter-rater reliability on the 5 double coded articles after discussing the disagreements.

Data analysis

For this synthesis, we completed two types of data analysis methods for studies including raw data which could be further refined. When studies included data only in graphs, we extracted the mean scores using WebPlotDigitizer. The WebPlotDigitizer program measured the distance of scores on a graph from the *x*- and *y*-axis based on the scale set for the graph (Rohatgi, 2020). For effect sizes, we calculated Cohen's *d* for studies if the means, standard deviations, and group sizes were provided using an effect size calculator for comparing groups with different sample sizes (Lenhard & Lenhard, 2016). When the needed information was not provided to calculate Cohen's *d*, we reported the provided effect sizes from the study (i.e., Cohen's *d*, Hedges' *g*, Tau-U, Eta Squared, and NAP).

We also calculated total number of minutes of instruction. To calculate total number of minutes of instruction, we multiplied the average minutes per session with the average number of sessions. If the studies included a range for the minutes per session or number of sessions, we first averaged the minimum and maximum, then proceeded with multiplying the minutes per session by the number of sessions.

Results

We identified 21 articles for the synthesis, and one of the articles included two separate studies (Hacker et al., 2019), resulting in 22 studies included in the synthesis. Only 10 studies reported race and ethnicity demographics. When reported, the proportion of African American/Black participants ranged from 0.00 to 0.49, Hispanic participants ranged from 0.08 to 1.00, White participants ranged from 0.00 to 0.85, and Asian/Pacific Islander participants ranged from 0.02 to 0.11. Other race and ethnicity categories reported included: Bosnian (0.17), Native American (0.02), Indonesian (0.11), Two or More Races (0.04 to 0.06), and Other (0.02). Table 2 provides a summary of additional study characteristics including study location, research design, study sample size, participant types, language status, grade level, age, implementer, group size and study quality.

Mathematics-writing instruction outcomes

Across the 22 studies, we examined mathematics-writing instruction based on student performance on mathematics measures, mathematics-writing measures, mathematics-vocabulary measures, and reading comprehension measures on mathematics topics.

Mathematics measures

A total of 13 studies included a mathematics test as an outcome measure. For these studies, six group design studies included the needed data to calculate Cohen's *d*, four group design studies included alternative data, and three single-case design studies included Tau-U and non-overlapping pairs (NAP).

Quantitative Among the studies measuring mathematics, six studies included the needed information to compare treatment and control groups based on significance testing and Cohen's *d*. For these six studies, measures were focused on algebra, geometry, fractions, and word-problems solving with four studies including standardized measures and two studies including researcher created measures. Significance testing was completed for 23 measures of general mathematics between treatment groups with instruction in mathematics writing compared control groups. Of these comparisons, 14 comparisons were reported as significant and nine were reported as not significant. For the significant comparisons, Cohen's *d* ranged between 0.03 and 1.88.

Table 2 Summary of study characteristics ($N=22$)

Characteristic		n	%
Study location			
	United States	19	86.4
	International	3	13.6
Research design			
	Randomized-control trial	9	40.9
	Single-case design	4	18.2
	Quasi-experimental design	9	40.9
Study N			
	<100	15	68.2
	100 to 999	6	27.3
	>1,000	1	4.5
MD (%)			
	<100	3	13.6
	100	8	36.4
	NR	11	50.0
Language status			
	ELL included	8	36.4
	Not reported	14	63.6
Grade level			
	1 to 2	3	13.6
	3 to 5	11[c]	50.0
	6 to 8	7[c]	31.8
	9 to 12	3	13.6
Age			
	6 to 7	2	9.1
	8 to 10	0	0.0
	11 to 12	0	0.0
	>13	4	18.2
	NR	16	72.7
Implementer			
	Educator	12	54.5
	Researcher	9	40.9
	NR	1	4.5
Group size			
	Individual	2[d]	9.1
	Small group[a]	5[d]	22.7
	Small group[b]	4[d]	18.2
	Whole class	11[d]	50.0
	NR	2	9.1
Study quality			
	<50%	3	13.6
	50% to 80%	13	59.1
	>80%	6	27.3

Table 2 (continued)

ELL, English language learners; MD, mathematics difficulty; NR, not reported

[a]Small group size between 2 and 8 students; [b]Size of small group not reported; [c]Two studies included in more than one category; [d]One study included in more than one category

For Cross et al. (2009), although significance testing was completed for one treatment group, reported with effects of significance testing with Cohen's *d*, significance testing was not completed for two additional treatment groups with exposure to mathematics-writing instruction. The author team reported using a researcher created measure focused on algebra.

Qualitative Four group design studies included alternative data. Two of these studies reported significance testing between treatment and control groups on measures of mathematics, but without the needed information to compare based on Cohen's *d* (Hacker et al., 2019; Swanson et al., 2019). These studies included 12 comparisons between treatment and control groups for measures of mathematics, with seven significant effects and five nonsignificant effects. Swanson et al. (2019) reported using standardized measures of word-problem solving or computational fluency but did not include the needed information to calculate for effect sizes. The standardized measures included the Comprehensive Math Achievement Test (CMAT), KeyMath, Test of Math Ability (TOMA-3), and Wechsler Individual Achievement Test: Arithmetic Computations (WIAT-R). Hacker et al. (2019) reported Hedges' *g* as 0.60 based on the standardized fractions measure *easyCBM Numbers and Operations.*

Two additional studies reported on measures of mathematics but did not include significance testing across the treatment and control groups. Blanton et al. (2019) reported that the treatment group grew at a significantly faster rate than the control group from pretest to posttest in Grade 3, with the treatment group showing a 21% advantage over the control group by the end of the school year on general mathematics. In Grades 4 and 5, the treatment group performed marginally significantly higher than the control group on general mathematics, with a 2% advantage over the control group. The author team reported using researcher-created measure focused on algebra. In Stoyle and Morris (2017), treatment and control groups were included, but not statistically compared. Treatment group one made greater gains growth from pretest ($M = 6.95$) to posttest ($M = 39.73$) than the control group from pretest ($M = 22.21$) to posttest ($M = 43.05$). Treatment group two also made greater gains from pretest ($M = 10.35$) to posttest ($M = 40.50$) than the control group. This study included a standardized measure in fractions from the Bridges mathematics program.

Single-case design Three single-case design studies included measures of mathematics. Two studies reported a Tau-U between 0.66 and 0.77 across all students and one study reported NAP across all students as 0.32 (Bundock et al., 2021; Hacker et al., 2019; Kong & Swanson, 2019). Two of the three used standardized measures in fractions and computational fluency; however, the name of the frac-

tion standardized measure was not reported. The computational fluency measure was AIMSweb Math Concepts and Applications (M-CAP). Two of the three also used researcher created measures focused rate of change or word-problem solving.

Mathematics-writing measures

A total of 11 studies included mathematics-writing measures. Of the studies with a mathematics-writing measure, five group design studies compared the effects of a treatment group to a control group with the needed information to calculate Cohen's *d*, three group design studies included alternative data, and two studies measured the difference in student performance from baseline to treatment conditions.

Quantitative Across the five studies to include the needed information to calculate Cohen's *d*, 23 comparisons of significance in relation to mathematics writing were completed, including measures of writing organization, mathematics vocabulary, and mathematics content in writing. Of these comparison, 14 comparisons were reported as significant and nine were reported as not significant. For the significant effects, Cohen's *d* ranged between 0.36 and 2.88. These studies all included researcher-created measures. Across the studies, the mathematics-writing measures focused on mathematics content including word-problem solving, fractions, and geometry.

Qualitative Of the four group-design studies to include alternative data, three included comparisons across groups and one included pretest to posttest comparisons. Across the three studies with comparisons across groups but without the needed information to calculate Cohen's *d*, nine comparisons of significance in relation to writing organization, mathematics vocabulary, and mathematics content in writing were completed, all of which were reported as significant. Chasanah et al. (2020) reported that the two treatment groups significantly outperformed the control group on a researcher-created measure of mathematics writing. Hacker et al. (2019) reported that the treatment group significantly outperformed the control group on researcher-created measures of mathematical reasoning ($g = 1.82$), number of argumentative elements ($g = 3.20$), and total words written ($g = 1.04$). Last, Uswatum and Mariani (2020) reported that the treatment group significantly outperformed the control group on completeness of mathematical communication ability, proportion of completeness, and written mathematical communication enhancement. It was not reported if this was on researcher created or standardized measures. The mathematics content for the measures included word-problem solving for Chasanah et al. (2020) and fractions for Hacker et al. (2019) and Uswatum and Mariani (2020).

One study used an alternative method for measuring growth on mathematics writing. Gearing and Hart (2019) the students demonstrated significant growth from pre- to posttest on their mathematics-writing scores. Student growth was

measured through a researcher-created measure with word-problem solving as the mathematics topic in the measure.

Single-case design Two studies measured the difference in student performance from baseline to treatment conditions. In Bundock et al. (2021), the students demonstrated a non-significant overall Tau-U of 0.33 from baseline to treatment. In Hacker et al. (2019), the students demonstrated non-overlap across all students in mathematics-writing reasoning (0.61), rhetorical elements (0.45), and total words written (0.14). These two studies used researcher-created measures. The measure in Bundock et al. (2021) included a mathematics focus of rate of change, and the measure in Hacker et al. (2019) focused on fraction content.

Other assessments

Two studies included alternative measures related to mathematics and mathematics writing.

Qualitative Yang and Lin (2012) measured reading comprehension of geometric proofs through a researcher-created measure. The students in the treatment group significantly outperformed the students in the control group for reading comprehension of geometric proofs for foundational knowledge, logical status, summary, and generality, but not for reading comprehension for proofs for application.

Single-case design Fore III et al. (2007) included a researcher-created measure of mathematics vocabulary. The students made significant growth from baseline to treatment with a 43% increase in performance from baseline to treatment (NAP = 0.78).

Outcomes based on study and instructional features

We also synthesized outcomes of studies with mathematics-writing instruction based on study and instructional features: Type of mathematics writing, instructional focus, and participant type.

Type of mathematics writing

For type of mathematics writing, we categorized outcomes of the studies by type of mathematics writing.

Exploratory writing Five studies focused only on exploratory writing during instruction. Two of the studies with exploratory writing completed significance testing comparing treatment and control groups with the needed data to calculate Cohen's *d*. They reported 10 significant comparisons for general mathematics (Cohen's $d = 0.06$ to 1.04) and eight non-significant comparisons for general mathematics (Moran et al., 2014; Swanson et al., 2014). Two additional studies reported significance testing to

compare treatment and control groups without the needed data to calculate Cohen's *d*. Swanson et al. (2019) reported six significant comparisons and five non-significant comparisons for general mathematics. Uswatun and Mariani (2020) reported three significant comparisons for mathematics writing. Last, one single-case study demonstrated growth on general mathematics from baseline to treatment (Tau-U = 0.66; Kong & Swanson, 2019).

Informative writing An additional six studies focused exclusively on informative writing during instruction. Of the six studies, three studies completed significance testing with the needed data to calculate Cohen's *d*. There were two significant comparisons for mathematics (Cohen's *d* = 0.03 to 0.06; Thayer & Giebelhaus, 2001) and five significant comparisons for mathematics writing (Cohen's *d* = 0.84 to 2.41; Hebert et al., 2019; Hughes & Lee, 2020). There was also one non-significant comparison for mathematics writing (Hughes et al., 2019). Three studies also completed alternative methods of analyzing results. Blanton et al. (2019) reported significant gains for the treatment group compared to the control group for Grade 3, but only marginally significant gains for Grade 4 or 5 on a measure of general mathematics. Gearing and Hart (2019) reported significant pretest to posttest gains on a measure of mathematics writing for the treatment group, but not for the control group. Using a single-cased design, Fore III et al. (2007) reported significant gains from baseline to treatment for mathematics vocabulary.

Argumentative writing Five studies focused only on argumentative mathematics writing during instruction. Of these studies, two studies completed significance testing with the needed data to calculate Cohen's *d*. There was one significant comparison for mathematics (Cohen's *d* = 0.82; Kiuhara et al., 2019) and three significant comparisons for mathematics writing (Cohen's *d* = 1.66 to 2.88; Kiuhara et al., 2019). Two additional studies reported comparisons between a treatment and control group for mathematics and mathematics writing without the needed data to calculate Cohen's *d*. This included one significant comparison for mathematics, three significant comparisons for mathematics-writing, four significant comparisons for reading comprehension of geometric proofs, and one non-significant comparison for reading comprehension of geometric proofs (Hacker et al., 2019; Yang & Lin, 2012). Last, one single-case design study reported students made growth on mathematics (NAP = 0.32) and mathematics writing for reasoning (NAP = 0.61), rhetorical elements (NAP = 0.45), and total words written (NAP = 0.14) from baseline to treatment (Hacker et al., 2019).

Mathematically-creative writing We coded for studies that used only mathematically-creative writing, but no studies included in this synthesis reported only using mathematically-creative writing.

Multiple types of writing The remaining six studies included multiple types of mathematics writing. Five of the six studies included argumentative and informative mathematics writing. For the studies with argumentative and informative mathematics

writing, three studies reported significance testing between a treatment and control group with the needed data to calculate Cohen's *d*. The studies reported eight significant comparisons for mathematics writing (Cohen's $d = 0.36$ to 0.59), two significant comparisons for mathematics (Cohen's $d = 0.37$ to 1.88), eight non-significant comparisons for mathematics writing, and one non-significant comparison for mathematics (Cohen et al., 2015; Cross, 2009; Gavin et al., 2013). Although Cross (2009) reported significance testing for one treatment group, they did not report significance testing for two additional treatment groups. Additionally, Stoyle and Morris (2017) did not report significance testing for two treatment groups but did report the treatment groups made greater gains than the control group from pretest to posttest on mathematics. Last, in Bundock et al. (2021), the students demonstrated a non-significant overall Tau-U of 0.33 from baseline to treatment on mathematics writing and a significant overall Tau-U of 0.77 from baseline to treatment on mathematics.

One of the six studies to include multiple types of mathematics writing was Chasanah et al. (2020) which included informative and creative writing. In this study, the two treatment groups significantly outperformed the control groups.

Instructional focus

Although the majority of the studies focused instruction on mathematics writing, several studies also focused instruction on mathematics content or mathematics vocabulary with mathematics writing as a component of the instruction.

Mathematics writing A total of 13 studies focused instruction on mathematics writing. Of these, six studies included comparisons between treatment and control groups with the needed data to calculate Cohen's *d*. These two studies reported two comparisons on mathematics where the treatment group significantly outperformed the control group (Cohen's $d = 0.37$ to 0.82; Cross, 2009; Kiuhara et al., 2019). Although Cross (2009) reported that one treatment group significantly outperformed the control, the authors did not significantly compare two additional treatment groups to the control group. For mathematics writing, the treatment groups significantly outperformed the control groups on 16 comparisons (Cohen's $d = 0.36$ to 2.88), but there was no difference between the treatment and the control groups for nine comparisons for mathematics writing (Cohen et al., 2015; Hebert et al., 2019; Hughes & Lee, 2020; Hughes et al., 2019; Kiuhara et al., 2019).

Five additional studies included treatment and control groups, but did not include the needed data to calculate Cohen's *d*. For mathematics writing, the treatment groups significantly outperformed the control on eight comparisons (Chasanah et al., 2020; Hacker et al., 2019; Uswatum & Mariani, 2020). For mathematics, the treatment group significantly outperformed the control for one comparison (Hacker et al., 2019). Two studies also reported significant gains from pretest to posttest for three comparisons of mathematics and mathematics writing for treatment groups (Gearing & Hart, 2019; Stoyle & Morris, 2017).

Two single-case design studies included a main instructional focus of mathematics writing. Bundock et al. (2021) reported significant gains for mathematics writing but not for general mathematics. Hacker et al. (2019) reported gains for mathematics writing and general mathematics.

Mathematics content Eight studies focused on mathematics content but used mathematics writing as a component of the instruction. A total of four studies included significance testing with the needed data to calculate Cohen's *d*. These studies included 22 comparisons of mathematics between treatment and control groups and no comparisons for mathematics writing. For the mathematics comparisons, 13 resulted in the treatment groups significantly outperforming the control group (Cohen's *d* = 0.03 to 1.88) and nine resulted in no significant difference (Gavin et al., 2013; Moran et al., 2014; Swanson et al., 2014; Thayer & Giebelhaus, 2001). Two additional studies reported significance testing on measures of mathematics between treatment and control groups without the needed data to calculate Cohen's *d* for 17 comparisons. For these comparisons, 11 were significant and six were not significant (Swanson et al., 2019; Yang & Lin, 2012). One additional study reported significant gains in mathematics from pretest to posttest in mathematics for Grade 3, but only marginally significant gains for Grades 4 and 5 (Blanton et al., 2019). And last, Kong and Swanson (2019) demonstrated growth on mathematics from baseline to treatment.

One final study included mathematics writing as a component of mathematics-vocabulary instruction. In the mathematics-vocabulary study, students made significant gains from baseline to treatment for mathematics vocabulary (Fore III et al., 2007).

Participant type

We also examined the effects based on participant type. We categorized the participant types as studies which identified all students in the sample as students with MD, studies which identified a portion of the students in the sample as students with MD, and studies which did not identify any students in the sample as students with MD.

Students with MD (100%) A total of eight studies identified all students in the sample as students with MD. Of these eight studies, three included significance testing between treatment and control groups with the needed data to calculate Cohen's *d*. There were 11 significant comparisons for mathematics (Cohen's *d* = 0.06 to 1.04) and eight non-significant comparisons for mathematics (Kiuhara et al., 2019; Moran et al., 2014; Swanson et al., 2014). Kiuhara et al. (2019) also reported one significant comparison for mathematics writing (Cohen's *d* = 2.88). Hacker et al. (2019) also identified all students in the sample as students with MD and compared a treatment group and control, but did not include the needed data to calculate Cohen's *d*. The treatment group significantly outperformed the control

for mathematics ($g = 0.60$) and for mathematics writing in relation to mathematical reasoning ($g = 1.82$), number of argumentative elements ($g = 3.20$), and total words written ($g = 1.04$). Three additional single-case design studies also identified all students in the sample as students with MD. For these studies, authors reported significant gains in mathematics vocabulary (Fore III et al., 2007) and positive gains in mathematics writing (Bundock et al., 2021; Hacker et al., 2019). For mathematics, Hacker et al. (2019) reported positive gains, but Bundock et al. (2021) reported non-significant gains.

Students with MD (< 100%) Three studies reported students with MD as a portion of the students included in the study sample. Gearing and Hart (2019) reported 4.7% of the students in the study were identified as students with MD. They did not disaggregate the results for students with MD, instead reported that the treatment group made significant gains on mathematics writing from pretest to posttest (Cohen's $d = 0.87$), but the control group did not make significant gains. Alternatively, Cohen et al. (2015) and Swanson et al. (2019) disaggregated the results of students with MD from the rest of the sample. In Cohen et al. (2015), 55% of the students were identified with MD. The students with MD made significant gains compared to the control group in mathematics writing in linking words, reasoning, formal vocabulary count, formal vocabulary terms used correctly, and complete sentences. They did not make significant gains compared to the control group for informal vocabulary content, informal vocabulary terms used correctly, and attempted mathematical writing. This differed from the students without MD who significantly outperformed the control group in mathematics writing for reasoning, formal vocabulary count, and formal vocabulary terms used correctly, but not for linking words, informal vocabulary content, informal vocabulary terms used correctly, attempt at mathematical writing, and complete sentences. Swanson et al. (2019) reported 33% of the sample included students with MD. The author team disaggregated the students with MD who were English language learners from the students with average mathematics performance who were English language learners. The students with MD in treatment group one did not significantly outperform the control group in any measure of mathematics, differing from the students without MD who significantly outperformed the control group on the TOMA-2 and CMAT. The students with MD in treatment group two significantly outperformed the control group for Key Math ($g = 0.23$), but not for the TOMA-2, CMAT, or WIAT-R. Alternatively, the students without MD in treatment group two significantly outperformed the control group for the TOMA-2, CMAT, and Key Math. The students with MD in treatment group three significantly outperformed the control group for the TOMA-2 and CMAT ($g = 0.22$), but not for the Key Math or WIAT-R, aligning with the students without MD.

Students without MD The remaining 11 studies did not identify students included in the studies as students with MD. Five of the studies with students without MD included significance testing between treatment and control groups with the needed data to calculate Cohen's *d.* For mathematics, the treatment group signifi-

cantly outperformed the control group for four comparisons (Cohen's $d = 0.03$ to 1.88) and did not significantly outperform the control group for one comparison (Cross et al., 2009; Gavin et al., 2013; Tayer & Giebellaus, 2001). Yet, Cross et al. (2009) did not report significance testing for two out of the three treatment groups compared to the control group. For mathematics writing, the treatment group significantly outperformed the control group for five comparisons (Cohen's $d = 0.84$ to 2.41) and did not significantly outperform the control group for one comparison (Hebert et al., 2019; Hughes & Lee, 2020; Hughes et al., 2019). The remaining studies provided information on treatment and control groups, but without the needed data to calculate Cohen's d. Chasanah et al. (2020) and Uswatun and Mariani (2020) included five significant comparisons between the treatment and control groups on mathematics writing. Yang and Lin (2012) included three significant comparisons and one non-significant comparison for reading comprehension of geometric proofs. Last, Blanton et al. (2019) and Stoyle and Morris (2017) reported significant gains in mathematics from pretest to posttest for treatment groups but not control groups. Blanton et al. (2019) reported significant gains from pretest to posttest for Grade 3 and marginally significant gains for Grade 4 and 5. Stoyle and Morris (2017) reported significant gains for two treatment groups from pretest to posttest.

Instructional methods

We examined methods in the 22 studies used to practice mathematics writing during mathematics-writing instruction (see Table 3). Across the studies, seven included educator *modeling* of mathematics writing. Studies more frequently included *discussion*, with 16 studies including *discussion*. Other embedded oral practices involved *think aloud* (2 studies), *educator questioning* (1 study), and *feedback* (3 studies).

Several studies included planning and organization methods related to mathematics writing. Eight studies featured *attack strategies* for mathematics writing. For example, Hughes et al. (2019) used PRISM-Check as an attack strategy. In PRISM-Check, students completed a six-step process for solving and responding in writing to the mathematics problem. Similarly, eight studies used *graphic organizers*, and five studies used *SRSD* as instructional methods. In Moran et al. (2014) students also *revised* their mathematics writing. In Cohen et al. (2015), the educators and students referenced an interactive *word wall*.

The structure of mathematics-writing instructional methods also ranged across studies. The practice of *journal writing* occurred in two studies with both paper journals and online journals (i.e., blogging). Alternatively, no studies included *letter writing*. In eight studies, students participated in *informative writing* tasks after solving mathematics problems. Similarly, students *paraphrased* mathematics concepts, procedures, and problems in five studies. Students also participated in *argumentative writing* to justify their mathematics problem solving in nine studies and responded to *formal mathematics-writing prompts* in three studies. In a single study, students *created problems* (Chasanah et al., 2020). Students also

Table 3 Study intervention components, methods, and results

Study	Design	Participants		MD[d] (%)	Total instruction time (min)	Group Size	Implementer	Instruction focus	MW Type	Mathematics content	Mathematics MW instructional components	Assessment type	Results	QI
		Treatment (n)	Control (n)											
Blanton et al. (2019)	RCT	1,087	1,079	N	3,240	Whole Class	Educator	MC	I	Algebra Readiness/ Algebra	Discussion; I (Solving mathematics problems)	Mathematics Test	[e]Students in the TXT grew at a significantly faster rate in Grade 3 from pretest ($M = 13.64\%$) to posttest ($M = 40.42\%$) than students in the CTL from pretest ($M = 14.90\%$) to posttest ($M = 27.31\%$), with a 21% advantage over CTL students by the end of the year. In the last two years, the TXT performed marginally significantly higher in Grades 4 ($M = 52.82\%$) and Grade 5 ($M = 64.67\%$), than students in the CTL in Grade 4 ($M = 38.84\%$) and Grade 5 ($M = 47.80\%$), with TXT demonstrating a 2% advantage above the CTL	0.63
Bundock et al. (2021)	SCD	4		Y (100)	210	Indiv.	Researcher	MW	A; I	Word-Problem Solving	A (Justifying mathematics); Attack strategy; Graphic organizer; I (Solving mathematics problems); Note taking	Mathematics and MW Test	[e]Writing Scores: Non-significant combined ESs for students' writing scores between the baseline and intervention (Tau-U=0.33) [e]Mathematics Scores: Significant combined ESs for mathematics scores from baseline to intervention (Tau-U=0.77)	0.82

Table 3 (continued)

Study	Design	Participants Treatment (n)	Control (n)	MD[1] (%)	Total instruction time (min)	Group Size	Implementer	Instruction focus	Instruc-tion MW Type	Mathematics content	MW instructional components	Assessment type	Results	QI
Chasanah et al. (2020)	QED; Pretest equivalence													
Treatment Group 1		100	57	N	NR	NR	Educator	MW	I; C	Word-Problem Solving	Creating Problems	MW Test	[c]TXT 1 (M = 80.92) significantly outperformed CTL (M = 66.84). Total possible points 100	0.38
Treatment Group 2		83					Educator				Creating Problems		[c]TXT 2 (M = 84.86) significantly outperformed CTL (M = 66.84). Total possible points 100	
Cohen et al. (2015)	QED; Pretest equivalence	193	191	Y (55)	NR	Whole Class	Educator	MW	A; I	Geometry/ Measurement	A (Justifying mathematics); Discussion; Graphic organizer; I (Solving mathematics problems; Word wall	MW Test	(+) The high-level students in the TXT performed significantly higher than the CTL for reasons (Cohen's d = 0.50), formal vocabulary count (Cohen's d = 0.49), and formal vocabulary terms used correctly (Cohen's d = 0.36). The high-level TXT and CTL did not significantly differ for linking words, informal vocabulary content, informal vocabulary terms used correctly, attempt at mathematical writing, and complete sentences	0.71

Table 3 (continued)

Study	Design	Participants Treatment (n)	Control (n)	MDᴵ (%)	Total instruction time (min)	Group Size	Implementer	Instruction focus	Instruc- MW Type	MW Mathematics content	Mathematics MW instructional components	Assessment type	Results	QI
													(+) The students with MD in the TXT performed significantly higher than the CTL students for linking words (Cohen's $d=0.45$), reasons (Cohen's $d=0.44$), formal vocabulary count (Cohen's $d=0.59$), formal vocabulary terms used correctly (Cohen's $d=0.42$), and complete sentences (Cohen's $d=0.25$) The students with MD in the TXT and CTL did not significantly differ for informal vocabulary content, informal vocabulary terms used correctly, and attempt at mathematical writing	
Cross (2009)	QED; Pretest equivalence													
Treatment Group 1		43	55	N	NR	Whole Class	Educator	MW	A; I	Algebra Readiness/Algebra	No MW	Mathematics Test	Authors did not statistically compare TXT 1 versus CTL	0.46
Treatment Group 2		51				Whole Class	Educator	MW	A; I	Algebra Readiness/Algebra	A (Justifying mathematics); I (Solving mathematics problems)		Authors did not statistically compare TXT 2 versus CTL	

Table 3 (continued)

Study	Design	Participants		MDᵈ (%)	Total instruction time (min)	Group Size	Implementer	Instruction focus	MW Type	Mathematics content	Mathematics MW instructional components	Assessment type	Results	QI
		Treatment (n)	Control (n)											
Treatment Group 3		62				Whole Class	Educator	MW	A; I	Algebra Readiness/Algebra	A (Justifying mathematics); Discussion; I (Solving mathematics problems)		(+) TXT 3 significantly outperformed the CTL (Cohen's d=0.37)	
Fore III et al. (2007)	SCD	6		Y (100)	200	NR	NR	MV	I	NR	Graphic organizer; Modeling; Note taking	MV Test	°On average student percent accuracy increased from baseline (M=63.75%) to treatment (M=90.67%) with a 42.23% increase (NAP=0.78)	0.86
Gavin et al. (2013)	QED; Pretest equivalence	186	174	N	NR	Whole Class	Educator	MC	A; I	Geometry; Measurement	A (Justifying mathematics); Defining vocabulary; Discussion; Graphic organizer; I (Solving mathematics problems); (Verbal) Paraphrasing	Mathematics Test	No significant difference between TXT and CTL for an overall mathematics scale (+) TXT significantly outperformed the CTL for a geometry measure (Cohen's d=1.88)	0.71
Gearing and Hart (2019)	RCT	23	27	Y (4.7)	NR	Whole Class	Researcher	MW	I	Operations; Word-Problem Solving	Discussion; I (Solving mathematics problems); Educator Questioning	MW Test	°Authors did not statistically compare TXT versus CTL. TXT significantly increase performance from pretest to posttest outperformed the CTL (Cohen's d=0.87). No significant difference from pretest to posttest for CTL	0.71
Hacker et al. (2019)														

Table 3 (continued)

Study	Design	Participants		MD¹ (%)	Total instruction time (min)	Group Size	Implementer	Instruction focus	MW Type	Mathematics content	Mathematics MW instructional components	Assessment type	Results	QI
		Treatment (n)	Control (n)											
Study 1	RCT	NR	NR	Y (100)	1,087.50	Small Group[b]	Educator	MW	A	Fractions	A (Justifying mathematics); Attack strategy; Discussion; Graphic organizer; Modeling; SRSD; Think aloud	Mathematics Test; MW Test	°TXT significantly outperformed CTL (g=0.60) °The TXT significantly outperformed the CTL in mathematical reasoning (g=1.82), number of argumentative elements (g=3.20), and total words written (g=1.04)	0.79
Study 2	SCD	34	NR	Y (100)	NR	Whole Class	Educator	MW	A	Fractions	A (Justifying mathematics); Attack strategy; Discussion; Graphic organizer; Modeling; SRSD; Think aloud	Mathematics Test; MW Test	°The average NAPs for all five classrooms for computational accuracy was 0.32 (range 0.13—0.69) °The average NAPs for all educators for mathematical reasoning was 0.61 (0.08—0.97); number of rhetorical elements was 0.45 (0.15—0.81); and for total words written was 0.14 (0.54—0.78)	0.77
Hebert et al. (2019)	RCT	29	32	N	360	Small Group[a]; Indiv.	Researcher	MW	I	Word-Problem Solving	Attack strategy; Feedback; I (Solving mathematics problems); Modeling; MW prompts;	MW Test	(+) For the total writing and mathematics score, the TXT significantly outperformed the CTL (Cohen's d=1.39) (+) For Percent Correct Mathematics Writing Sequences, the TXT significantly outperformed CTL (Cohen's d=0.84)	0.79

Table 3 (continued)

Study	Design	Participants Treatment (n)	Control (n)	MD[1] (%)	Total instruction time (min)	Group Size	Implementer	Instruction focus	MW Type	Mathematics content	MW instructional components	Assessment type	Results	QI
Hughes and Lee (2020)	QED; Pretest equivalence	13	12	N	495	Whole Class	Researcher	MW	1	Fractions; Decimals; Word-Problem Solving	Attack strategy; Discussion; I (Solving mathematics problems); Modeling; SRSD	MW Test	(+) For writing quality, the TXT significantly outperformed the CTL for the first word problem (Cohen's $d=1.27$), the second word problem (Cohen's $d=1.30$), and the two-word problems combined (Cohen's $d=1.76$) (+) For writing parts, the TXT significantly outperformed the CTL for the first word problem (Cohen's $d=2.17$), the second word problem (Cohen's $d=1.55$), and the two-word problems combined (Cohen's $d=2.41$) (+) For writing language, the TXT significantly outperformed the CTL for the first word problem (Cohen's $d=1.82$), the second word problem (Cohen's $d=1.10$), and the two-word problems combined (Cohen's $d=1.58$)	0.75
Hughes et al. (2019)	RCT	18	9	N	270	Whole Class	Researcher	MW	1	Word-Problem Solving	Attack strategy; Discussion; I (Solving mathematics problems); Modeling; SRSD	MW Test	No significant difference between TXT and CTL	0.79

Table 3 (continued)

Study	Design	Participants Treatment (n)	Control (n)	MD[1] (%)	Total instruction time (min)	Group Size	Implementer	Instruction focus	Instruc-MW Type	Mathematics MW content	Mathematics MW instructional components	Assessment type	Results	QI
Kiuhara et al. (2019)	RCT	28	31	Y (100)	1,395	Small Group[b]	Educator	MW	A	Fractions	Attack strategy; A (Justifying mathematics); Defining vocabulary; Discussion; Graphic organizer; Modeling; Note taking; SRSD	Mathematics Test; MW Test	(+) TXT significantly outperformed the CTL (Cohen's $d = 0.82$) (+) For total rhetorical elements, the TXT significantly outperformed the CTL (Cohen's $d = 2.88$) (+) For quality of mathematics reasoning, the TXT significantly outperformed the CTL (Cohen's $d = 2.71$) (+) For total words written, the TXT significantly outperformed the CTL (Cohen's $d = 1.66$)	0.96
Kong and Swanson (2019)	SCD	9		Y (100)	630	Small Group[a]	Researcher	MC	E	Word-Problem Solving	Attack strategy; Paraphrasing	Mathematics Test	°Overall, the 9 students, the Tau-U from baseline to treatment was 0.66 with a 95% confidence interval between 0.33 and 1.0	0.82
Moran et al. (2014) *Treatment Group 1*	RCT	16	13	Y (100)	550	Small Group[a]	Researcher	MC	E	Word-Problem Solving	Discussion; Paraphrasing; Revising	Mathematics Test	No significant difference between TXT 1 and CTL for Key Math problem solving, CMAT, and TOMA-2	0.67

Table 3 (continued)

Study	Design	Participants Treatment (n)	Control (n)	MD[d] (%)	Total instruction time (min)	Group Size	Implementer	Instruction focus	MW Type	Mathematics content	MW instructional components	Assessment type	Results	QI
Treatment Group 2		19			550	Small Group[a]	Researcher	MC	E	Word-Problem Solving	Discussion; Paraphrasing; Revising		(+) TXT 2 significantly outperformed CTL for Key Math (Cohen's $d=0.46$), CMAT (Cohen's $d=0.74$), and TOMA-2 (Cohen's $d=0.65$)	
Treatment Group 3		24			550	Small Group[a]	Researcher	MC	E	Word-Problem Solving	Discussion; Paraphrasing; Revising		(+) TXT 3 significantly outperformed CTL for Key Math (Cohen's $d=0.21$), CMAT (Cohen's $d=1.04$), and TOMA-2 (Cohen's $d=0.53$)	
Stoyle and Morris (2017)	QED; Pretests and random assignment to treatment groups			N										
Treatment Group 1		60	19		90	Whole Class	Educator	MW	A; I	Fractions	A (Justifying mathematics); I (Solving mathematics problems); Journal writing (blog)	Mathematics Test	[c]Authors did not statistically compare TXT 1 versus CTL. Authors noted TXT 1 had greater growth from pretest ($M=6.95$) to posttest ($M=39.73$) than CTL from pretest ($M=22.21$) to posttest ($M=43.05$)	0.58
Treatment Group 2		55			90	Whole Class; Small Group[b]	Educator	MW	A; I	Fractions	A (Justifying mathematics); Discussion; I (Solving mathematics problems); Journal writing		[c]Authors did not statistically compare TXT 2 versus CTL. Authors noted TXT 2 had greater growth from pretest ($M=10.35$) to posttest ($M=40.50$) than control	
Swanson et al. (2019)														

Table 3 (continued)

Study	Design	Participants		MD[d] (%)	Total instruction time (min)	Group Size	Implementer	Instruction focus	MW Type	Mathematics content	Mathematics MW instructional components	Assessment type	Results	QI
		Treatment (n)	Control (n)											
Treatment Group 1	RCT	30	45	Y (33)	600	Small Group[b]	Researcher	MC	E	Word-Problem Solving	Discussion; Paraphrasing; Feedback	Mathematics Test	[e] No significant difference between TXT 1 and CTL for TOMA-2, CMAT, Key Math, and WIAT-R	0.92
Treatment Group 2		29			600	Small Group[b]	Researcher	MC	E	Word-Problem Solving	Discussion; Paraphrasing; Feedback		[e] TXT 2 significantly outperformed the CTL for TOMA-2, CMAT, and Key Math [e] No significant difference between ELL TXT 2 and CTL for WIAT-R [e] Monolingual English-speaking students in TXT 2 significantly outperformed the CTL for WIAT-R	
Treatment Group 3		38			600	Small Group[b]	Researcher	MC	E	Word-Problem Solving	Discussion; Paraphrasing; Feedback		[e] TXT 3 significantly outperformed the CTL for TOMA-2, CMAT, and Key Math [e] No significant difference between all TXT 3 and CTL group for WIAT-R	
Swanson et al. (2014)														
Treatment Group 1	RCT	17	20	Y (100)	600	Small Group[b]	Researcher	MC	E	Word-Problem Solving	Discussion; Paraphrasing; Feedback	Mathematics Test	No significant difference between TXT 1 and CTL for CMAT, Key Math, and WIAT-R (+) TXT 1 significantly outperformed CTL for the Schema Assessment (Cohen's $d = 0.57$)	0.88

Table 3 (continued)

Study	Design	Participants Treatment (n)	Control (n)	MD[d] (%)	Total instruction time (min)	Group Size	Implementer	Instruction focus	MW Type	Mathematics MW content	MW instructional components	Assessment type	Results	QI
Treatment Group 2		20			600	Small Group[b]	Researcher	MC	E	Word-Problem Solving	Discussion; Paraphrasing; Feedback		(+) There was a significant post-test advantage for TXT 2 over CTL for CMAT and KeyMath combined (Cohen's $d=0.06$) No significant difference between TXT 2 and CTL groups for WIAT_R (+) TXT 2 significantly outperformed the CTL for the Schema Assessment (Cohen's $d=0.64$)	
Treatment Group 3		25			600	Small Group[b]	Researcher	MC	E	Word-Problem Solving	Discussion; Paraphrasing; Feedback		No significant post-test advantage for TXT 2 over CTL for CMAT, KeyMath, and WIAT_R (+) TXT 3 significantly outperformed the CTL for the Schema Assessment (Cohen's $d=0.55$)	
Thayer and Giebelhaus (2001)	QED; Pre-test and random assignment	18	20	N	NR	Whole Class	NR	MC	I	Geometry	Journal writing; MW prompts	Mathematics Test	(+) For the chapter 11 mathematics test, the TXT significantly outperformed the CTL (Cohen's $d=0.06$). For the chapter 12 mathematics test, the TXT group significantly outperformed the CTL group (Cohen's $d=0.03$)	0.71

Table 3 (continued)

Study	Design	Participants Treatment (n)	Control (n)	MD[d] (%)	Total instruction time (min)	Group Size	Implementer	Instruction focus	MW Type	Mathematics content	Mathematics MW instructional components	Assessment type	Results	QI
Uswatun and Mariani (2020)	QED; Pretest equivalence	40	40	N	NR	Small Group[a]	Educator	MW	E	Fraction	Discussion	MW Test	[c]The TXT significantly outperformed the CTL for completeness of written mathematical communication ability, proportion of completeness, and written mathematical communication enhancement	0.38
Yang and Lin (2012)	QED; Pretest non-equivalence—posttest scores adjusted for non-equivalence	345	338	N	225	Whole Class	Educator	MC	A	Geometry	A (Justifying mathematics); Discussion; Graphic organizer; MW prompts; Paraphrasing	Reading Test	[c]The TXT significantly outperformed the CTL for reading comprehension of geometric proofs for basic knowledge, logical status, summary, and generality / No significant difference between all TXT and CTL for reading comprehension of geometric proofs for application	0.71

A, argumentative writing; C. creative writing; CMAT, comprehensive math achievement test; CTL, control group; E, exploratory writing; I, informative writing; Indiv., individual; MC, mathematics content; MD, mathematics difficulty; MW, mathematics writing; MV, mathematics vocabulary; N, no; NAP, non-overlap for all pairs; NR, not reported; RCT, randomized-control trial; SCD, single-case design; TOMA-2, test of math ability, version 2; TXT, treatment group; QED, quasi-experimental design; QI, quality indicator checklist; WIAT-R, Wechsler individual achievement test: arithmetic computations; Y, yes; (+) = significant effects

[a]Small group size between 2 and 8 students; [b]Size of small group not reported; [c]Cohen's d between treatment and control could not be calculated; [d]Percent MD for entire sample

used mathematics writing for *defining vocabulary* in two studies and *note taking* in three studies.

Instructional methods for students with MD

Finally, we reviewed the instructional methods for practicing mathematics writing during mathematics-writing instruction in eight studies which identified all students in the sample as students with MD. Four studies included educator *modeling* in mathematics writing. In five studies, students participated in *discussions* related to mathematics writing, two of which students also engaged in *think aloud*. Additionally, for one study which included discussion, the educator also provided *feedback* to the students (Swanson et al., 2014).

These studies also included instructional methods related to organization and planning. Of the studies with instructional methods related to organization and planning, five studies included an *attack strategy*, and five studies included *graphic organizers*. Three of the studies with an *attack strategy* and *graphic organizers* used the *SRSD* framework. Additionally, in one study students *revised* their mathematics writing (Moran et al., 2014).

The studies supporting students with MD through mathematics-writing instruction also varied in mathematics-writing formats. Although no studies included *journal writing*, *letter writing*, or *formal mathematics-writing prompts*, studies included *argumentative writing*, *note taking*, *informative writing*, *paraphrasing*, and *defining vocabulary*. Students justified their reasoning through *argumentative writing* in three studies and participated in *informative writing* tasks after solving mathematics problems in three studies. Similarly, in three studies students with MD also commonly participated in *note taking*. Students with MD also participated in *paraphrasing* in three studies and *defined vocabulary* with only one study (Kiuhara et al., 2019).

Discussion

In this synthesis, we reviewed studies with mathematics-writing instruction to review outcomes and methods of mathematics-writing instruction for students with and without MD. First, we asked about student outcomes in mathematics writing and mathematics for students who participated in mathematics-writing instruction. Then, we investigated whether outcomes varied based on study and instructional features. Third, we examined which methods authors used to practice mathematics writing. And fourth, we determined which methods authors used to practice mathematics writing for studies with only students with MD. By examining the outcomes and methods of mathematics-writing instructional practices, considerations for future research and practice can be identified.

Is mathematics-writing instruction helpful?

Across Grades 1 to 12, authors generally reported positive academic outcomes related to instruction with mathematics writing. Like the previous mathematics-writing synthesis (Powell et al., 2017), the outcome measures for studies included a wide variety of measures. For the studies with treatment and control groups with mathematics measures, the author teams reported treatment groups significantly outperformed the control groups more frequently than they reported no significant difference. For the studies using pre- to posttesting or baseline to intervention differences, the author teams also reported growth in mathematics outcomes.

For the studies with treatment and control groups measuring mathematics-writing measures, authors reported the treatment groups significantly outperformed the control groups more frequently than they reported no significant difference. For the studies using pre- to posttest or baseline to intervention differences, the author teams reported significant mathematics-writing gains on one study (Gearing & Heart, 2019), growth in mathematics-writing on the second study (Hacker et al., 2019), and no significant gains on the third study (Bundock et al., 2021). One study measured student performance through mathematics vocabulary, with students demonstrating significant gains in mathematics vocabulary (Fore III et al., 2007). Last, one author team measured reading comprehension of geometric proofs, with students demonstrating mixed outcomes (Yang & Lin, 2012).

The generally positive results suggest that student mathematics and mathematics-writing outcomes may improve after exposure to instruction with mathematics writing. The positive trend aligns with the previous synthesis on mathematics writing which indicated promising results for mathematics-writing practices such as journal writing and organized mathematics writing in the classroom (Powell et al., 2017). Results of this synthesis demonstrated the use of mathematics-writing instruction may support student outcomes for mathematics and mathematics writing on high-stakes assessments. Yet, with the high variability of type of measure, mathematics content, and methods of analysis, more research is needed to determine efficacy of mathematics-writing instruction on student outcomes in mathematics and mathematics writing.

Do the study and instructional features impact mathematics-writing efficacy?

Type of mathematics writing

First, we examined the type of mathematics writing used within mathematics-writing instruction. Studies either focused on one type of mathematics writing or multiple types of mathematics writing. Of the studies focused on one type of mathematics writing, author teams most frequently included only informative writing (6 studies), aligning with previous research (Powell et al., 2017). Unlike previous research, the synthesized results indicated that nearly as many studies focused exclusively on exploratory writing (5 studies). For informative writing studies, overall studies demonstrated positive mathematics and mathematics-writing outcomes. The exploratory

writing studies demonstrated higher rates of non-significant outcomes across studies, especially with for measures of mathematics. Slightly fewer studies exclusively included argumentative writing (5 studies), with positive mathematics and mathematics-writing outcomes. Alternatively, zero studies included only creative writing. Generally, we noted improved outcomes for students when they participated in mathematics-writing instruction including informative, exploratory, argumentative, or creative writing.

Six more studies included multiple types of mathematics writing within instruction. Five of the studies included both argumentative and informative writing. These studies demonstrated about an even spread of significant and non-significant outcomes for mathematics and mathematics writing when comparing the treatment and control groups. One study focused on informative and creative mathematics writing in which the students in the treatment group significantly outperformed the students in the control group on mathematics writing. Across the six studies to include multiple types of mathematics writing, the minimal significant results may suggest that involving multiple types of mathematics writing within instruction may restrict student outcomes. Yet, due to the limited number of studies included, no conclusions can be drawn about the limitations of including multiple types of mathematics writing in instruction.

Instructional focus

In addition to types of mathematics writing, we examined efficacy based on instructional focus. Just over half of the studies (13 studies) focused on mathematics writing as the primary area of instruction. For the studies focused on mathematics writing, author teams of frequently reported positive outcomes for either mathematics or mathematics-writing outcomes. Such trends align with educator perspectives that the use of mathematics-writing effectively supports students to learn mathematics and communicate mathematically (Powell et al., 2021).

The remaining eight studies focused on mathematics or mathematics vocabulary, including mathematics writing as one component of overall mathematics instruction. Of the mathematics-focused studies, author teams reported a mix of significant and non-significant outcomes for mathematics. None of these studies reported mathematics-writing outcomes. The inclusion of mathematics writing may support student mathematics outcomes; however, conclusions cannot be drawn based on how mathematics writing impacted mathematics outcomes due to the multi-component nature of the interventions. Last, for the one study focused on mathematics vocabulary, students demonstrated significant gains on mathematics vocabulary. Therefore, including mathematics writing as a component of mathematics-vocabulary instruction may be of value, but more research could help determine if mathematics writing was an active component.

Participant type

We also examined the outcomes for studies with a full range of students and studies with only students with MD in the treatment group(s). Eight studies

exclusively included students with MD, three studies reported including students with MD as a portion of the study sample, and 11 studies did not report including students with MD. For the studies with only students with MD, students most frequently demonstrated significant performance on mathematics and less frequently demonstrated significant performance on mathematics writing. The studies to include a portion of the sample as students with MD (between 4.7 and 55%) only disaggregated the students with MD for two studies (Cohen et al., 2015; Swanson et al., 2019). For these two studies, students significantly outperformed the students in the control group only for some measures of mathematics and mathematics writing. The remaining 11 studies did not report the number of students with MD. For these studies, students in the treatment groups frequently outperformed the students in the control groups on mathematics and mathematics writing. While a limited number of studies with mixed measures and findings produced these results, it does indicate that students with MD can benefit from mathematics-writing instruction. Such trends suggest that mathematics-writing may be used effectively to support mathematics and mathematics-writing outcomes for students with MD in addition to a full range of students.

How do educators teach mathematics writing?

Across studies, authors used a wide variety of methods for instruction in mathematics writing. The two methods used most often included *discussion* and explaining problem solving with *informative writing*. The high rates of inclusion of *discussion* aligned with reports that educators who taught mathematics-writing engaged students in guided practice, provided corrective feedback, practices which relate to the use of educator and student discussions (Powell et al., 2021). The trends indicate the importance of *discussion* within mathematics-writing instruction. Furthermore, the common use of responding to problem-solving prompts reflects the value placed on *informative writing* by researchers and educators (Powell et al., 2021).

All other strategies occurred in 10 or fewer studies. Between 5 and 10 studies included: *graphic organizers*, *argumentative writing* to justify mathematics problem solving, *attack strategies*, *modeling*, *paraphrasing*, and *SRSD*. Under five studies included the following instructional practices: *note taking*, *defining vocabulary*, *journal writing*, *formal mathematics-writing prompts*, *feedback*, *think aloud*, *educator question*, *revised mathematics writing*, *word wall*, and *create problems*. Additionally, no studies involved *letter writing*, *respond to discussion*, or *drawing*. Several of the methods align with practices commonly used to support mathematics instruction, writing instruction, or both mathematics and writing instruction. For example, the use of *attack strategies* frequently occurs in mathematics instruction to support word-problem solving (Fuchs et al., 2021; Powell et al., 2020). Alternatively, the use of SRSD is frequently used as an explicit writing instruction framework for general writing skills (Graham, 2005). Other strategies, such as *graphic organizers*, *vocabulary instruction*, and *feedback* support students across the areas

of mathematics and writing (Alghamdi et al., 2020; Graham et al., 2013; Kim et al., 2021; Powell & Driver, 2015). The high variety of methods used across studies suggest that few methods have been established as best practices for mathematics-writing instruction; instead, authors tend to rely on methods commonly used in mathematics instruction or writing instruction.

How do educators teach mathematics writing to students with MD?

Many of the same methods used in the studies with a full range of students were also used in the studies with only students with MD in the treatment group(s). As described, only eight studies focused exclusively on students with MD. Five studies included the methods of *discussion*, *attack strategy*, and *graphic organizers* and four studies included *modeling*. Three studies included *argumentative writing*, *informative writing*, *SRSD*, *paraphrasing*, and *note taking*. Fewer than three studies included *think aloud*, *feedback*, *revising*, and *defining vocabulary*. No studies with a sample of only students with MD included *journal writing*, *letter writing*, and *formal mathematics-writing prompts*. Across the studies focused exclusively on students with MD and on a wider range of students, we identified *discussion* as one of the most common methods for instruction for mathematics writing. This may indicate the value of including discussion as a tool for supporting mathematics-writing development for a full range of students as well as students with MD. Other methods frequently used for students with MD also overlapped with methods used in mathematics and writing instruction, such as *attack strategies* and *graphic organizers* (Alghamdi et al., 2020; Hebert et al., 2018; Powell et al., 2020). The results indicates that, similar to the methods used across all studies, methods for mathematics and writing instruction also overlap with those used for students with MD in mathematics-writing instruction. More research could better identify the efficacy of the methods for directly supporting students with MD in mathematics writing.

Limitations

Before providing considerations for research and practice, we note several limitations. First, because of the variability in research designs and limited quantitative data provided by authors, we did not have access to enough data to provide effect sizes for each study. Although we reported standardized effect sizes for Cohen's *d* for 10 group design studies, the remaining 8 group design studies are qualitatively compared because they did not include the needed information to calculate Cohen's *d*. Additionally, four single-case design studies either reported Tau-U or NAP. If all studies could be compared with consistent effect sizes, this would provide easier comparisons across studies and allow for a meta-analytic approach to the data. Similarly, the lack of any standardized measures in mathematics writing or consistent use of the same mathematics measures prohibits comparison of results across studies. Second, the results are limited by publication bias. In this synthesis we only have the significance testing included which had been reported by the author teams.

Some author teams did not report significance testing (Cross et al., 2009) and others may have excluded measures which did not demonstrate significant differences. Third, even though 22 studies appear adequate, it is important to note the studies occurred across Grades 1 to 12. Therefore, only a handful of studies focused on any one grade level, so the number of studies is often inadequate to draw strong conclusions about mathematics-writing instructional practices. Similarly, the number of studies focused exclusively on students with MD was less than 10. Taken together, future research is warranted to increase the study numbers and sample sizes which would allow for more robust conclusions about mathematics writing.

Considerations for research and practice

Our results demonstrate a high need for future research on mathematics-writing instruction. Researchers should primarily consider that despite the positive results indicating mathematics writing supports students, more high-quality studies are needed to strengthen the knowledge base on mathematics-writing instruction. By conducting studies with clear instructional plans, formal mathematics and mathematics-writing measures, and a rigorous data analysis plan, a stronger understanding of the efficacy of mathematics-writing instruction and practices can be determined. The second consideration for research involves the development of mathematics-writing instruction for students with MD. Students with MD tend to experience more difficulty than their peers without MD on mathematics-writing measures (Arsenault et al., 2022; Hughes et al., 2020), but few studies focused on students with MD. More rigorous research could identify how to intensify mathematics-writing outcomes for students with MD.

Considering the positive outcomes and the increased prevalence of mathematics writing (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010; Powell & Hebert, 2022), we recommend that practitioners provide instruction in mathematics writing to both classes with a full range of students and intervention groups including students with MD. When providing mathematics-writing instruction, practitioners should consider instruction on one type of mathematics writing, especially informative mathematics writing, due to the frequent use and high rates of positive outcomes for students practicing informative mathematics writing. They should also consider using best practices emphasized in the literature on writing and mathematics, such as *discussions*, *graphic organizers*, or *attack strategies* within mathematics-writing instruction.

Conclusion

We examined mathematics writing in the elementary and secondary grades. Many studies reported positive outcomes in mathematics and mathematics writing. Such positive outcomes varied across study and instructional features, such as types of mathematics writing, instructional focus, and participant type. For studies with a full range of students and studies primarily supporting students with MD, instructional

methods aligned with those used in mathematics and writing instruction (Alghamdi et al., 2020; Graham et al., 2013; Kim et al., 2021; Powell & Driver, 2015). While many studies in this synthesis demonstrated improved mathematics writing, results also indicate the high need for future rigorous research on mathematics-writing instruction to provide high-quality analysis of mathematics-writing instruction for students with and without MD.

References

*Included in synthesis

Alghamdi, A., Jitendra, A. K., & Lein, A. E. (2020). Teaching students with mathematic disabilities to solve multiplication and division word problems: The role of schema-based instruction. *ZDM Mathematics, 52*, 125–137. https://doi.org/10.1007/s11858-019-01078-0

Alvi, E., & Nausheen, M. (2019). Examining grade 9 students' engagement in mathematical problem-solving (MPS) when working as individuals and in a small group Settings. *Bulletin of Education and Research, 41*(1), 163–184.

Arsenault, T. L., & Powell, S. R. (2022). Word-problem performance differences by schema: A comparison of students with and without mathematics difficulty. *Learning Disabilities Research & Practice*. https://doi.org/10.1111/ldrp.12273

Arsenault, T. L., Powell, S. R., Hebert, M. A., King, S. G., Lin, X., & Lang, D. (2022). Mathematics-writing profiles for students with mathematics difficulty. *Reading and Writing*. https://doi.org/10.1007/s11145-022-10375-2

Baxter, J. A., Woodward, J., & Olson, D. (2005). Writing in mathematics: An alternative form of communication for academically low-achieving students. *Learning Disabilities Research & Practice, 20*(2), 119–135. https://doi.org/10.1111/j.1540-5826.2005.00127.x

*Blanton, M., Stroud, R., Stephens, A., Gardiner, A. M., Stylianou, D. A., Knuth, E., Isler-Baykal, I., & Strachota, S. (2019). Does early algebra matter? The effectiveness of an early algebra intervention in Grades 3 to 5. *American Educational Research Journal, 56*(5), 1930–1972. https://doi.org/10.3102/0002831219832301

Broto, J. A., & Greer, R. D. (2014). The effects of functional writing contingencies on second graders' writing and responding accurately to mathematical algorithms. *Behavioral Development Bulletin, 19*(1), 7–18. https://doi.org/10.1037/h0100568

*Bundock, K., Hawken, L. S., Kiuhara, S. A., O'Keeffe, B. V., O'Neill, R. E., & Cummings, M. B. (2021). Teaching rate of change and problem solving to high school students with high incidence disabilities at tier 3. *Learning Disability Quarterly, 44*(1), 35–49. https://doi.org/10.1177/0731948719887341

Casa, T. M., Firmender, J. M., Cahill, J., Cardetti, F., Choppin, J. M., Cohen, J., Cole, S., Colonnese, M. W., Copley, J., DiCicco, M., Dieckmann, J., Dorl, J., Gavin, M. K., Hebert, M. A., Karp, K. S., LaBella, E., Moschkovich, J. N., Moylan, K., Olinghouse, N. G., Powell, S. R., Price, E., Pugalee, D. K., Rupp Fulwiler, B., Sheffield, L. J., & Zawodniak, R. (2016). *Types of and purposes for elementary mathematical writing: Task force recommendations.* https://mathwriting.education.uconn.edu/wp-content/uploads/sites/1454/2016/04/Types_of_and_Purposes_for_Elementary_Mathematical_Writing_for_Web-2.pdf

Casler-Failing, S. L. (2018). Integrating LEGO robotics into a 5th grade cross curricular unit to promote the development of narrative writing skills. In *Proceedings of the interdisciplinary STEM teaching and learning conference, 2*(6), 36–42. https://doi.org/10.20429/stem.2018.020106

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards mathematics*. http://corestandards.org/

*Chasanah, C., Riyadi, R., & Usodo, B. (2020). The effectiveness of learning models on written mathematical communication skills viewed from students' cognitive styles. *European Journal of Educational Research, 9*(3), 979–994. https://doi.org/10.12973/eu-jer.9.3.979

*Cohen, J. A., Casa, T. M., Miller, H. C., & Firmender, J. M. (2015). Characteristics of second graders' mathematical writing. *School Science and Mathematics, 115*(7), 344–355. https://doi.org/10.1111/ssm.12138

Çontay, E. G., & Duatepe-Paksu, A. (2018). The effect of journal writing on achievement and geometry self-efficacy of 8th grade students. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education, 12*(2), 167–198.

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education, 36*(4), 220–234. https://doi.org/10.1177/0741932514557271

Correnti, R., Matsumura, L. C., Hamilton, L., & Wang, E. (2013). Assessing students' skills at writing analytically in response to texts. *The Elementary School Journal, 114*(2), 142–177. https://doi.org/10.1086/671936

*Cross, D. I. (2009). Creating optimal mathematics learning environments: Combining Argumentation and writing to enhance achievement. *International Journal of Science and Mathematics Education, 7*(5), 905–930. https://doi.org/10.1007/s10763-008-9144-9

Dunston, P. J., & Tyminski, A. M. (2013). What's the big deal about vocabulary? *Mathematics Teaching in the Middle School, 19*(1), 38–45. https://doi.org/10.5951/mathteacmiddscho.19.1.0038

Fello, S. E., & Paquette, K. R. (2009). Talking and writing in the classroom. *Mathematics Teaching in the Middle School, 14*(7), 410–414. https://doi.org/10.5951/MTMS.14.7.0410

*Fore, C., III., Boon, R. T., & Lowrie, K. (2007). Vocabulary instruction for middle school students with learning disabilities: A comparison of two instructional models. *Learning Disabilities: A Contemporary Journal, 5*(2), 49–73.

Fuchs, L. S., Seethaler, P. M., Sterba, S. K., Craddock, C., Fuchs, D., Compton, D. L., Geary, D. C., & Changas, P. (2021). Closing the word-problem achievement gap in first grade: Schema-based word-problem intervention with embedded language comprehension instruction. *Journal of Educational Psychology, 113*(1), 86–103. https://doi.org/10.1037/edu0000467

Fuentes, S. Q. (2011). A square becomes a regular octagon: An authentic experience in proof writing. *Australian Senior Mathematics Journal, 25*(2), 14–24. https://doi.org/10.3316/informit.593231834342458

*Gavin, M. K., Casa, T. M., Firmender, J. M., & Carroll, S. R. (2013). The impact of advanced geometry and measurement curriculum units on the mathematics achievement of first-grade students. *Gifted Child Quarterly, 57*(2), 71–84. https://doi.org/10.1177/0016986213479564

*Gearing, N. V., & Hart, L. C. (2019). The impact of adding written discourse to six year olds' mathematics explanations within a problem-based learning unit. *European Journal of STEM Education*. https://doi.org/10.20897/ejsteme/3952

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology, 104*(1), 206–223. https://doi.org/10.1037/a0025398

Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing, 26*(1), 1–15. https://doi.org/10.1007/s11145-012-9395-2

Graham, S., Harris, K. R., & Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology, 30*(2), 207–241. https://doi.org/10.1016/j.cedpsych.2004.08.001

*Hacker, D. J., Kiuhara, S. A., & Levin, J. R. (2019). A metacognitive intervention for teaching fractions to students with or at-risk for learning disabilities in mathematics. *ZDM Mathematics Education, 51*(4), 601–612. https://doi.org/10.1007/s11858-019-01040-0

Hebert, M., Bohaty, J. J., Nelson, J. R., & Roehling, J. V. (2018). Writing informational text using provided information and text structures: An intervention for upper elementary struggling writers. *Reading and Writing, 31*(9), 2165–2190. https://doi.org/10.1007/s11145-018-9841-x

Hebert, M. A., & Powell, S. R. (2016). Examining fourth-grade mathematics writing: Features of organization, mathematics vocabulary, and mathematical representations. *Reading and Writing, 29*(7), 1511–1537. https://doi.org/10.1007/s11145-016-9649-5

*Hebert, M. A., Powell, S. R., Bohaty, J. J., & Roehling, J. (2019). Piloting a mathematics writing intervention with late elementary students at-risk for learning difficulties. *Learning Disabilities Research and Practice, 34*(3), 144–157. https://doi.org/10.1111/ldrp.12202

Hecht, S. A., & Vagi, K. J. (2010). Sources of group and individual differences in emerging fraction skills. *Journal of Educational Psychology, 102*(4), 843–859. https://doi.org/10.1037/a0019824

*Hughes, E. M., & Lee, J.-Y. (2020). Effects of a mathematical writing intervention on middle school students' performance. *Reading & Writing Quarterly, 36*(2), 1–17. https://doi.org/10.1080/10573569.2019.1677537

*Hughes, E. M., Lee, J. Y., Cook, M. J., & Riccomini, P. J. (2019). Exploratory study of a self-regulation mathematical writing strategy: Proof-of-concept. *Learning Disabilities: A Contemporary Journal, 17*(2), 185–203.

Hughes, E. M., Riccomini, P. J., & Lee, J.-Y. (2020). Investigating written expressions of mathematical reasoning for students with learning disabilities. *The Journal of Mathematical Behavior, 58*, 100775. https://doi.org/10.1016/j.jmathb.2020.100775

Kim, Y.-S.G., Yang, D., Reyes, M., & Connor, C. (2021). Writing instruction improves students' writing skills differentially depending on focal instruction and children: A meta-analysis for primary grade students. *Educational Research Review, 34*, 100408. https://doi.org/10.1016/j.edurev.2021.100408

*Kiuhara, S. A., Rouse, A. G., Dai, T., Witzel, B. S., Morphy, P., & Unker, B. (2019). Constructing written arguments to develop fraction knowledge. *Journal of Educational Psychology, 112*(3), 584–607. https://doi.org/10.1037/edu0000391

*Kong, J. E., & Swanson, H. L. (2019). The effects of a paraphrasing intervention on word problem-solving accuracy of English learners at risk of mathematic disabilities. *Learning Disability Quarterly, 42*(2), 92–104. https://doi.org/10.1177/0731948718806659

Kosko, K. W., & Norton, A. (2012). Relationships between the process standards: Process elicited through letter writing between preservice teachers and high school mathematics students. *School Science and Mathematics, 112*(6), 340–348. https://doi.org/10.1111/j.1949-8594.2012.00151.x

Kosko, K. W., & Zimmerman, B. S. (2019). Emergence of argument in children's mathematical writing. *Journal of Early Childhood Literacy, 19*(1), 82–106. https://doi.org/10.1177/1468798417712065

Lenhard, W., & Lenhard, A. (2016). *Computation of effect sizes*. Psychometrica. https://doi.org/10.13140/RG.2.2.17823.92329

Levenberg, İ, & Shaham, C. (2014). Formulation of word problems in geometry by gifted pupils. *Journal for the Education of Gifted Young Scientists, 2*(2), 28–40. https://doi.org/10.17478/jeysg.201429019

Monroe, E. E., & Panchyshyn, R. (1995). Vocabulary considerations for teaching mathematics. *Childhood Education, 72*(2), 80–83. https://doi.org/10.1080/00094056.1996.10521849

*Moran, A. S., Swanson, H. L., Gerber, M. M., & Fung, W. (2014). The effects of paraphrasing interventions on problem-solving accuracy for children at risk for math disabilities. *Learning Disabilities Research & Practice, 29*(3), 97–105. https://doi.org/10.1111/ldrp.12035

Namkung, J. M., Hebert, M., Powell, S. R., Hoins, M., Bricko, N., & Torchia, M. (2019). Comparing and validating four methods for scoring mathematics writing. *Reading & Writing Quarterly, 36*(2), 1–19. https://doi.org/10.1080/10573569.2019.1700858

National Council of Teachers of Mathematics. (2014). *Principles to actions*. National Council of Teachers of Mathematics. https://www.nctm.org

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

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

Powell, S. R., Berry, K. A., & Benz, S. A. (2020). Analyzing the word-problem performance and strategies of students experiencing mathematics difficulty. *The Journal of Mathematical Behavior, 58*, 100759. https://doi.org/10.1016/j.jmathb.2020.100759

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

Powell, S. R., & Hebert, M. A. (2016). Influence of writing ability and computation skill on mathematics writing. *The Elementary School Journal, 117*(2), 310–335. https://doi.org/10.1086/688887

Powell, S. R., & Hebert, M. A. (2022). Thinking beyond symbols: Writing and reading in mathematics. In Z. A. Philippakos & S. Graham (Eds.), *Writing and reading connections: Bridging research and classroom practice*. Guilford.

Powell, S. R., Hebert, M. A., Cohen, J. A., Casa, T. M., & Firmender, J. M. (2017). A synthesis of mathematics writing: Assessments, interventions, and surveys. *Journal of Writing Research, 8*(3), 493–526. https://doi.org/10.17239/jowr-2017.08.03.04

Powell, S. R., Hebert, M. A., & Hughes, E. M. (2021). How educators use mathematics writing in the classroom: A national survey of mathematics educators. *Reading and Writing: An Interdisciplinary Journal, 34*(2), 417–447. https://doi.org/10.1007/s11145-020-10076-8

Rohatgi, A. (2020). *Webplotdigitizer: Web based tool to extract data from plots, images, and maps* (Version 4.4) [Computer software]. https://automeris.io/WebPlotDigitizer/index.html

Stonewater, J. K. (2002). The mathematics writer's checklist: The development of a preliminary assessment tool for writing in mathematics. *School Science and Mathematics, 102*(7), 324–334. https://doi.org/10.1111/j.1949-8594.2002.tb18216.x

*Stoyle, K. L., & Morris, B. J. (2017). Blogging mathematics: Using technology to support mathematical explanations for learning fractions. *Computers & Education, 111*, 114–127. https://doi.org/10.1016/j.compedu.2017.04.007

*Swanson, H. L., Kong, J. E., Moran, A. S., & Orosco, M. J. (2019). Paraphrasing interventions and problem-solving accuracy: Do generative procedures help English language learners with math difficulties? *Learning Disabilities Research & Practice, 34*(2), 68–84. https://doi.org/10.1111/ldrp.12194

*Swanson, H. L., Moran, A., Lussier, C., & Fung, W. (2014). The effect of explicit and direct generative strategy training and working memory on word problem-solving accuracy in children at risk for math difficulties. *Learning Disability Quarterly, 37*(2), 111–123. https://doi.org/10.1177/0731948713507264

Tan, T., & Garces-Bacsal, R. M. (2016). The effect of journal writing on mathematics achievement among high-ability students in Singapore. *Gifted and Talented International, 28*(1–2), 173–184. https://doi.org/10.1080/15332276.2013.11678412

*Thayer, J., & Giebelhaus, C. R. (2001). Integrating writing to improve math achievement. *Mid-Western Educational Researcher, 14*(2), 11–16.

*Uswatun, K., & Mariani, S. (2020). Comparison between generative learning and discovery learning in improving written mathematical communication Ability. *International Journal of Instruction, 13*(3), 729–744.

Utami, C. T. P., Mardiyana, S., & Triyanto, S. (2019). Profile of students' mathematical representation ability in solving geometry problems. *IOP Conference Series: Earth and Environmental Science, 243*(1), 012123. https://doi.org/10.1088/1755-1315/243/1/012123

van Drie, J., van Boxtel, C., Jaspers, J., & Kanselaar, G. (2005). Effects of representational guidance on domain specific reasoning in CSCL. *Computers in Human Behavior, 21*(4), 575–602. https://doi.org/10.1016/j.chb.2004.10.024

*Yang, K.-L., & Lin, F.-L. (2012). Effects of reading-oriented tasks on students' reading comprehension of geometry proof. *Mathematics Education Research Journal, 24*(2), 215–238. https://doi.org/10.1007/s13394-012-0039-2