Piloting a Mathematics-Writing Intervention with Late Elementary Students At-Risk for Learning Difficulties

Michael A. Hebert

University of Nebraska—Lincoln

Sarah R. Powell

The University of Texas at Austin

Janet J. Bohaty

University of Nebraska—Lincoln

Julia Roehling

Georgia Southern University

Paper should be cited as:
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 & Practice, 34*(3), 144-157. https://doi.org/10.1111/ldrp.12202

Author Note

Michael A. Hebert, School of Education, University of California, Irvine.

This research was supported in part by the Institute of Education Sciences, U.S. Department of Education, through award R324B130005 to the University of Nebraska. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

Correspondence concerning this article should be addressed to Michael Hebert, 401 East Peltason Drive, Suite 3100, Irvine, CA 92697. E-mail: mhebert1@uci.edu.

Abstract

High-stakes mathematics assessments require students to write about mathematics, although research suggests students exhibit limited proficiency on such assessments. Students with LD may have difficulties with writing, mathematics, or both. Researchers employed an intervention for teaching students how to organize mathematics writing (MW). Researchers randomly assigned participants (*n* = 61) in grades 3-5 to receive instruction in MW or information writing. Students receiving MW outperformed control students on a researcher-developed measure of MW (*d* = 1.05). Component assessment revealed MW students improved in writing organization (*d* = 1.49) but not in mathematics content (*d* = 0.11 *ns*). Results also indicated MW students outperformed control on percentage of correct MW sequences (*d* = 0.82). Future directions for MW intervention development are discussed.

Keywords: mathematics, writing, mathematics-writing, learning disabilities

Piloting a Mathematics-Writing Intervention
With Late Elementary Students At-Risk for Learning Difficulties

With the development of the Common Core State Standards in the United States (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010), the mathematics expectations for elementary students have increased beyond previously published state standards (Porter, McMaken, Hwang, & Yang, 2011). Along with changes to the mathematics content taught in classrooms, the introduction of the Common Core forced changes to high-stakes mathematics assessments. Current assessments do not only require students to solve problems with whole and rational numbers, judge inequalities, determine lines of symmetry, solve equations, and perform measurement conversions, among other tasks; current assessments also require students to write about mathematical processes or judgments.

After reviewing newer assessments associated with the Common Core (e.g., Partnership of Readiness for College and Career [PARCC] and SmarterBalanced), it is clear that demonstration of mathematics competency now involves disciplinary writing in the area of mathematics. We call this *mathematics writing* (MW). Because of this high-stakes focus on MW, it is important for teachers to provide instruction to help students write about mathematics effectively. In this study, we piloted a mathematics-writing (MW) intervention for at-risk students in the late elementary grades to examine the promise of such an intervention. In our introduction, we describe current MW assessments and the scoring of such assessments. We then highlight MW intervention efforts and the outcomes of such efforts. Finally, we provide a purpose for this pilot study and outline our research questions.

Mathematics-Writing Assessments

On the 2017 version of a mathematics performance task from Smarter Balanced (www.smarterbalanced.com), students answered mathematics questions using both numbers and writing. For example, one prompt asked students to “explain the steps you used to figure this out.” Another prompt requested students “explain why or why not” and “use the information shown in your explanation.” A 3-point scoring rubric (i.e., full and complete understanding, partial understanding, or limited understanding) was utilized to score student answers. On another assessment associated with the Common Core (www.parcconline.org), assessment prompts asked fourth-grade students to “explain the error than Jian made…,” “explain why Shaun’s reasoning is incorrect,” and “identify the incorrect reasoning in Christy’s statement…explain how Christy can correct her reasoning.” In a manner somewhat similar to the Smarter Balanced scoring, student MW was scored against 3- and 4-point rubrics.

With these examples from large-scale, high-stakes assessments, we note that, when students were prompted to write within mathematics, the majority of items prompted students to provide written explanations of their own mathematics work, or of the work of others. Scoring was holistic, with room for interpretation by the scorers. Unfortunately for teachers, the individual MW scores of students have been aggregated into overall mathematics scores. Therefore, it is impossible to determine how the MW of students differed from mathematics performance on questions that did not require writing. It is also impossible to gather data about the MW strengths and weaknesses of individual students. To inform instructional practices, it is necessary to understand how mathematics content and writing organization influence the scoring of MW.

In a recent synthesis of MW, Powell, Hebert, Cohen, Casa, and Firmender (2017) identified 18 studies with empirical data about MW assessments, but only three of the studies assessed the MW of students exclusively in the elementary grades. Across studies, teachers often used mathematics journals as the method for collection of student MW samples. Students most often participated in MW by constructing mathematical explanations, and the scoring of MW assessments varied widely. At second grade, Cohen, Miller, Casa, and Firmender (2015) scored MW according to writing features, mathematics features, and mathematics reasoning. That is, students received writing points for including linking words and complete sentences. Students also received mathematics points for using formal vocabulary terms and informal vocabulary terms correctly. Finally, Cohen et al. (2015) scored students' MW explanations according to a rubric where a score of 2 represented strong understanding of the mathematics, a score of 1 indicated limited understanding of the mathematics, and a score of 0 meant a response was incorrect, irrelevant, or incoherent. Additionally, at second grade, Kostos and Shin (2010) scored MW explanations according to a rubric with three categories (i.e., mathematical knowledge, strategic knowledge, and explanation). Kostos and Shin used a 0 to 4 scale within each category, with 4 demonstrating competence within the category. At fifth grade, Evens and Houssart (2004) scored MW by category (i.e., nothing written, wrong, restated, examples provided, or some degree of justification).

Across grade levels, students demonstrated limited proficiency with MW (Powell et al., 2017). For example, Barlow and Drake (2008) determined that, when asked to write a mathematics word problem, only 4% of students provided a satisfactory writing sample, and only 2% provided a correct and extended word problem. Jigyel and Afamasaga-Fuata'i (2007) asked fourth-grade students to compare fractions and write explanations about the comparisons; only 29% of students provided correct explanations. Similarly, Kasmer and Kim (2012) noted only 38% of middle schoolers provided sophisticated MW samples.

To better understand the driving forces behind stronger and weaker MW samples, Powell

& Hebert (2016) designed two MW measures for fourth-grade students. One of the measures activated knowledge about whole numbers through a multi-step word problem, and the other measure assessed fraction concepts. First, the authors learned that both narrative writing and mathematics computation acted as significant predictors for MW scores. Second, students wrote fewer words in their MW and included fewer introductory statements, conclusions, paragraphs, and transition words than in their narrative writing samples (Hebert & Powell, 2016). Interestingly, when comparing the MW between the two MW measures, students wrote fewer words in response to the fraction prompt, yet used more symbols within the MW about fractions. From this research, we learned that (a) there is wide variability in the MW scores of students, (b) students approach MW differently from narrative writing, and (c) MW differs based on mathematics content.

In sum, prior research in the elementary grades about MW assessments utilized different methods for scoring MW: rubric scoring (Kostos & Shin, 2010), awarding points for features of writing or mathematics (Powell & Hebert, 2016; Cohen et al., 2015), or scoring by category (Evens & Houssart, 2004). Each of these scoring methods shared similarities with scoring of general writing samples (Koutsoftas & Gray, 2012; Olinghouse & Wilson, 2013). We did not, however, locate any MW assessments in which student work was scored according to writing curriculum-based measurement (CBM) recommendations, such as percentage of correct word sequences, which is often described as an effective method for scoring the writing of students with or at-risk for learning difficulties (McMaster, Du, Parker, & Pinto, 2011; Ritchey & Coker, 2014).

Intervention Efforts for MW

Unfortunately, we are aware of no data from high-stakes assessments that is currently available about the MW scores of students in the late elementary grades. The more informal efforts to understand how students write about mathematics (e.g., Hebert & Powell, 2016; Cohen et al., 2015; Kostos & Shin, 2010) provide detail about how students write within mathematics, and it is clear that many students struggle with MW. To improve MW knowledge, researchers have conducted several investigations. While the majority of MW instruction has been conducted at the middle and high school levels, a few investigations have focused on the elementary grades (Powell et al., 2017).

At second grade, Kostos and Shin (2010) evaluated the mathematics journal writing of 16 second-grade students. During the first week of the project, a teacher provided three modeling examples about constructing mathematics written explanations in journals. During the next four weeks of the project, the teacher taught three mini-lessons about incorporating mathematics vocabulary into written explanations, identifying key numbers and terms in the question to better understand a mathematics problem, and providing a step-by-step explanation for how a problem was solved. In all, students participated in 16 independent journal-writing experiences over a period of 5 weeks. A pre- and post-test, in which students completed a mathematics pattern problem and then provided a written explanation of the work, demonstrated growth across the study's duration. That is, the average score at pretest was 7.25, and this score increased to 10.00 at post-test. In the Kostos and Shin study, all students participated in the journal writing, with no students who were not participating in MW acting as a control for maturation effects.

Similarly, at second grade, Cohen et al. (2015) carried out a 12-week study in which second-grade students participated in geometry and measurement lessons incorporating MW activities (n = 193), or acted as a comparison condition (i.e., no MW experiences; n = 191). Students participated in MW approximately every three days by responding to prompts encouraging students to “think deeply.” Students’ MW involved independently writing explanations to fictional characters. Teachers did not provide explicit MW instruction; rather, twice during the study, teachers encouraged students to analyze excellent and substandard MW examples to identify the characteristics of effective MW. At post-test, students in the MW condition demonstrated superior use of formal and informal mathematics vocabulary, and used more expert mathematical reasons in their MW, than students who did not practice MW.

In both studies (Cohen et al., 2015; Kostos & Shin, 2010), students received an introduction to MW, and then received MW practice opportunities. The teacher-led instruction or teacher-guided discussions, combined with student-level practice, led to increased MW at post-test. Both of these studies worked with second-grade students, and we were unable to identify any third-, fourth-, or fifth-grade MW studies with embedded instruction. As the majority of MW assessment questions begin appearing on late elementary mathematics assessments, it is important to know how late-elementary students respond to MW instruction, and whether such instruction increases MW scores.

In the previous MW intervention efforts, researchers worked with students in general education. Perhaps for this reason, authors did not rely on explicit instructional practices more often utilized with students with or at-risk for learning difficulties, such as teacher modeling followed by guided practice (Hughes, Morris, Therrien, & Benson, 2017). For example, Kostos and Shin (2010) incorporated a few examples of teacher modeling, but followed this modeling by asking students to complete independent MW practice. Cohen et al. (2015) did not provide explicit modeling from the teacher, but engaged students in periodic discussions with exemplar and non-exemplar examples of MW. Students then participated in independent MW. Neither intervention provided students with guided practice opportunities in which the teacher and students worked together on MW prompts. Furthermore, neither intervention provided consistent opportunities for teacher modeling or teacher-led discussions. For the majority of the time, students independently worked on MW prompts without a lot of feedback from the teacher.

Although not explicitly stated by the author teams, the student samples of Cohen et al. (2015) and Kostos and Shin (2010) could have included some students with or at-risk for learning difficulties. In previous research, students with learning difficulties have exhibited lower mathematics performance than students without learning difficulties (Swanson, Jerman, & Zheng, 2008), with students with learning difficulties in both mathematics and reading having the lowest scores (Cirino et al., 2015; Vukovic & Siegel, 2010). Related to MW, students with learning difficulties in mathematics also have exhibited difficulty with reading (Murphy, Mazzocco, Hanich, & Early, 2007), phonological processing (Landerl, Fussenegger, Moll, & Willburger, 2009), language (Cirino et al., 2015), and verbal memory (Andersson, 2010), which may impact their ability to write in mathematics.

Similarly, writing is a challenge for students with learning disabilities (Troia, 2006; Gillespie & Graham, 2014), due to difficulties with phonological skills, spelling and handwriting difficulties, and working memory constraints (Hebert, Kearns, Hayes, Bazis, & Cooper, 2018). Moreover, while 15% of students without disabilities scored below basic on the most recent writing assessment of the National Assessment of Educational Progress, 60% of students with disabilities scored below basic, and only 5% scored at or above proficient (National Center for Education Statistics, 2011). Because of the variety of challenges students may face with MW, it is necessary to conduct MW research with students with or at-risk for learning disabilities.

Purpose and Research Questions

To learn the impact of MW instruction provided to students with or at-risk for learning difficulties, we conducted a pilot study to understand whether brief instruction and multiple MW practice opportunities dispersed across several weeks improved the MW of students. Specifically, we asked the following research questions: (1) Does an MW intervention show promise of effectiveness for improving MW scores of writing organization and mathematical content? (2) Does a MW intervention show promise for improving students' MW curriculum-based measurement (CBM) scores of MW (i.e., percent correct MW sequences)?

For research question 2, we developed a scoring system for percent correct MW sequences. In traditional scoring of Correct Writing Sequences, symbols and numbers that are not spelled out are not included in the count (e.g., Powell-Smith & Shin, 2004). Naturally, MW necessitates the use of symbols and numbers, so ignoring them is non-tenable in MW activities.

Method

We conducted a randomized-control trial to examine the effectiveness of an MW intervention compared to an informational text writing comparison condition. We describe both conditions later in the Method section. The informational text writing condition was used as a strong counterfactual to the MW condition to ensure effects were not attributed simply to writing instruction but to MW instruction. This study was conducted at a university reading center using two cohorts. The first cohort of students completed the study during the summer, while elementary school was not in session, and the second cohort of students completed the study in the fall by participating in an after-school program. Study procedures and instruction were otherwise the same for both cohorts.

Participants

Fourth- and fifth-grade struggling readers who attended a university reading center were eligible for participation in the study. To attend in the reading center, students had to score one or more grade levels behind their peers on reading measures used by the reading center. Because the focus of the reading center is on reading and writing skills, we did not have any measures of mathematics performance. However, students who struggle with reading often also struggle with writing (Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008), and reading disabilities are also often comorbid with mathematics disabilities (Fletcher, 2005; Fuchs & Fuchs, 2002), which justified the inclusion of such students in this study.

Overall, we included 66 students in the study. We randomly assigned participants to one of two conditions: (1) MW or (2) informational text writing. The informational text writing condition acted as a comparison for the MW intervention and allowed us to control for intervention time spent with a tutor. Prior to the intervention, but following random assignment, five students (7.5%) decided not to participate in the study, citing scheduling conflicts with summer vacations. All five students had been randomly assigned to the MW condition. It was not feasible to redo the randomization procedures because tutors had already been assigned and schedules had been confirmed. Therefore, we decided to conduct the study with the differential attrition as a limitation.

After attrition, 61 fourth- and fifth-grade students participated in this study (38 fourth- and 23 fifth-grade students). The participants included 35 boys and 26 girls. Table 1 displays student demographics by condition. Due to the aforementioned attrition, the authors statistically compared the two treatment groups on demographics, basic reading skills and reading comprehension cluster scores on the Woodcock Reading Mastery 3rd edition (WRMT-III), and the Essay Composition subtest of the Weschler Individualized Achievement Test, third edition (WIAT-III; Psychological Corporation, 2009). We noted no statistically significant differences between groups on free or reduced lunch ($\chi^2_{(1)} = 0.04$, $p = .85$), IEP status ($\chi^2_{(1)} = 2.59$, $p = .11$), or ethnicity ($\chi^2_{(4)} = 3.02$, $p = .55$). Boys and girls, however, were not distributed proportionally among the MW and informational text writing conditions ($\chi^2_{(1)} = 5.11$, $p = .02$). We identified no statistical difference on scores for the WRMT-III and the WIAT-III.

Measures

Research assistants (RAs) administered the study-specific pre- and post-test measures to students in both conditions; the RAs were often blind to condition, although we did not specifically blind them. The first and third authors trained the RAs to administer all measures.

Woodcock Reading Mastery Test, 3rd edition (screener). We obtained students' scores for subtests of the WRMT-III for the Word Identification, Word Attack, and Passage Comprehension from the reading center. These tests were used to screen participants for the reading center eligibility, and hence for eligibility for the current study. As mentioned in the participant section, we used these measures to test for potential differences in reading ability across the groups. Publishers report the internal reliability of the measures as .94 (Word Identification), .94 (Word Attack), and .87 (Passage Comprehension), respectively, for 4th grade, and as .91, .92, and .86, respectively, for 5th grade.

Essay composition. We administered the WIAT-III Essay Composition subtest, a norm-referenced writing measure, at pretest (to examine potential differences in writing skill between the conditions) and post-test (to measure potential distal effects of the intervention). We selected this measure because it was closer to the types of non-fiction writing in the intervention conditions than other norm-referenced measures involving narrative or story writing. Students were given 10 min to write about their favorite game and write three reasons why that game was their favorite. We scored Essay Composition utilizing the scoring protocol for the WIAT-III. Students earned a maximum or 20 points as a raw score for their essay (0 to 2 points for the introduction, 0 to 2 points for a conclusion, 0 to 5 points for paragraphs, 0 to 5 points for transition words, 0 to 3 points for reasons why they like the game, and 0 to 3 points for an elaboration of each reason). Internal consistency reliability, as reported by Lichtenberger and Breaux (2010), is .86 for fourth grade and .87 for fifth grade.

Mathematics writing. At pre- and post-test, students also completed an MW assessment developed by the authors and validated in previous studies (see Hebert & Powell, 2016; Powell & Hebert, 2016). In the prompt, students see a multi-step word problem completed by a pseudo-student, "Sam," who solved the word problem in four steps. Step A involved multiplication of single-digit numbers, but this step is unnecessary because the information is within the word problem. In step B, Sam added two-digit numbers ($20 + $20), but step B should not have been completed. The word problem indicated a total of only $20, so Sam should not have added an additional $20. Step C involved addition of monetary values. This procedure was correct, but Sam made a regrouping mistake in the ones place. Step D required subtraction. Sam used the incorrect minuend (from step B) and the incorrect subtrahend (from step C). The student also procedurally lined up the numbers incorrectly for the subtraction problem. Thus, Sam arrived at an incorrect answer. To administer the prompt, the examiner read the prompt aloud while the students looked over Sam's work. Students then wrote for 10 min about Sam's mistakes and how to solve the problem correctly. We scored the MW measure in two different ways.

Organization and content. First, we scored each student's MW measure for organization and content. For the organization score, we developed a scoring system for writing organization based on the scoring rubric for WIAT-III Essay Composition. Similar to the WIAT-III Essay Composition subtest, students earned 0 to 2 points for their introduction, 0 to 2 points for a conclusion, 0 to 5 points for the number of paragraphs, and 0 to 5 points for each novel transition word. Maximum score for writing organization was 14. The content score was loosely based on the WIAT-III scoring, but instead of scoring for reasons and elaborations, we scored for (a) mathematics errors identified, (b) elaborations about the errors, and (c) inclusion of the corrected answers or explanations of how the answer could be corrected. Students could score points in these categories for each step of Sam's work. In step A, students earned up to 3 points for identifying the mathematics was correct, identifying the step was unnecessary, and explaining why it was unnecessary. We scored step B similarly with a 3-point maximum. In step C, students earned up to 4 points for identifying the operation was correct, identifying the answer was incorrect, elaborating about the mistake (i.e., regrouping error), and correcting the answer. In step D, students earned up to 9 points for identifying incorrect minuend and subtrahend, providing correct minuend and subtrahend, identifying incorrect place value setup, elaborating about place value, correcting place value, identifying incorrect subtraction, elaborating on incorrect subtraction, providing correct subtraction, and correcting the answer (i.e., writing the correct answer). Please note we did not explicitly ask students to provide the correct answer; we prompted students to "write about how you would solve the problem correctly." The maximum score for mathematics content was 33 points. Coefficient alpha (Cronbach's α) for this sample was .76. Every assessment was scored by two independent scorers. Two research assistants scored all of the writing samples. Inter-rater reliability, calculated as the percentage of agreement, was 97% based on item-by-item agreement.

Percentage of correct mathematics writing sequences. We also scored MW by adapting CBM scoring recommendations from general writing. For this study, we scored using percentage of correct MW sequences (%CMWS), which allowed us to examine the writing conventions around the specific incorporation of mathematical terms within written text. We adapted

%CMWS from the “percentage of correct word sequences” (%CWS) often used in scoring writing CBM (McMaster & Espin, 2007). The primary adaptation of %CWS was the inclusion of numbers and symbols in the scoring system. In many CBMs of writing, numerals and symbols are either not counted (i.e., ignored) or treated as incorrect, although exceptions are sometimes made for dates or money (e.g., Breaux, 2010). As such, we did not identify these rules as appropriate for scoring MW because students often include numbers, symbols, and equations within their MW. Therefore, it was necessary to develop a CBM of MW designed to score for conventions around the inclusion of numbers, symbols, and equations in writing.

With scoring correct word sequences, each writing unit (i.e., word) is scored according to whether two adjacent units are acceptable within written content. Scorers mark correct word sequences with a carat (^). Examples of correct word sequences include dog^ate or followed^by. Both tree°big and bikking°grass are examples of incorrect word sequences because of syntactical, grammar, or spelling errors. To score %CMWS, we examined every two adjacent writing units to determine whether the units were grammatically, syntactically, semantically, or contextually accurate within the context of the sentence. We developed special rules for scoring %CMWS involving two adjacent units involving (a) a word *and* number or symbol, and (b) a number or symbol *and* a number or symbol. The list of rules included guidelines for use of mathematical symbols to replace words, placement of symbols within the text, mixing of numbers and words within an equation, etc. (See Appendix A for the list of rules with examples.)

We calculated %CMWS scores by determining the number of Correct MW Sequences (CMWS) and Incorrect MW Sequences (IMWS). Then, %CMWS was calculated by dividing the total of CMWS by the sum of CMWS and IMWS. All writing samples were scored by two raters and then checked for reliability using point-by-point agreement. Interrater reliability was .94. We resolved discrepancies through discussion.

Materials

Due to the nature of the experiment comparing two interventions outside of a school setting, the study included materials and instruction for both the intervention (i.e., MW) and comparison (i.e., informational text writing) conditions.

Mathematics-writing intervention. The second author created MW exercises using previously released items from the National Assessment of Educational Progress (NAEP). In the spring of 2016, we downloaded all fourth-grade mathematics questions from the previous two administrations (i.e., 2013 and 2011) of the NAEP. We created 12 two-item exercise sets from word problems with an open response or word problems that could be rewritten to encourage an open response. We also targeted NAEP questions in which a common error that students make (e.g., regrouping error, multiplication error) could be featured within the worked example. Table 2 shows the content, domain, and difficulty level of the 12 NAEP items selected for the 12 exercise sets. Each set was designed for use within a 30-min lesson.

Each lesson's exercise set featured one NAEP prompt that was gently altered from the first exercise to the second exercise. Figure 1 shows an example of a two-exercise set. In the first exercise, the student was shown the work of a pseudo-student (Sam) who made a mistake when solving the problem. The MW directions asked the student to write about the mistakes Sam made and how to solve the problem correctly. The second exercise in each set was a variation of the first, but required the participants to complete a mathematics problem and then write about how they solved the problem. For lessons 7 through 12, the second exercise also prompted the student to use pertinent mathematics vocabulary in their writing. For example, if the mathematics problem involved multiplication, the second exercise might include words and definitions for terms such as *multiplication*, *each*, *regroup*, and *expect*. Student-friendly mathematics definitions were included for each vocabulary term.

Procedure

Following the pretest assessments, we randomly assigned participants to conditions. In both conditions, RAs taught the students one-on-one or in small groups of two, based primarily on scheduling. Students in both conditions received instruction in twelve 30-min lessons, controlling for instructional time across groups. RAs taught both the treatment and comparison conditions.

Mathematics-writing intervention. Students in the MW intervention completed twelve 30-min lessons. Each lesson was conducted in two parts. In the first exercise of the lesson, the tutor and student examined a word problem completed by a pseudo-student (Sam). Sam always made mistakes in the MW. The tutor and student identified Sam's mistakes, and determined how to fix the mistakes, and then the student wrote about how they would help Sam. The tutors taught the students to write using a multi-paragraph format that included writing an introduction, step-by-step instructions on how to fix each mistake made by Sam, and a conclusion. In the second exercise of the lesson, the tutor monitored as the student solved a similar problem and then wrote about how they solved the problem.

We designed the MW intervention for tutors to use a flexible instructional protocol involving modeling, guided practice, and independent practice. Lesson 1 used modeling only; the instructor completed both exercises (including the writing), while the student(s) observed. To keep the learning active, the tutor prompted the students to write everything the tutor wrote in the student workbook, including all of the steps for the mathematics problem. The tutor taught the student(s) to follow the following sequence:

For Exercise 1 of each lesson involving the pseudo-student: (1) Read the mathematics problem. (2) Examine the steps made by the pseudo-student (Sam) and try to identify mistakes. (3) Work the problem to correct the mistakes. (4) Write about the process, using a multi-paragraph format including an introduction, body, and conclusion. In the introduction, describe the mathematics problem and what the pseudo-student (Sam) was supposed to do, in terms of steps. Next, preview (write) the steps Sam completed correctly, and the steps in which Sam made mistakes. In the body, write about how to help Sam. If Sam did a step correctly, indicate that you should praise Sam by stating that this step was correct. If Sam did a step incorrectly, indicate how you would explain where Sam went wrong and what needs to be done instead. In the conclusion, write about the steps that Sam completed incorrectly, and briefly state how you helped Sam (e.g., "In summary, Sam made a mistake in regrouping. I explained the steps of the problem to Sam and helped him regroup correctly").

For Exercise 2 involving an unworked problem: (1) Read the vocabulary terms and definition that you are asked to use for the MW exercise. [Note: Step 1 was only included in lessons 7-12. All other lessons began with step 2.] (2) Read the mathematics problem. (3) Complete the mathematics problem. (4) Write about how you completed the mathematics problem using an introduction, body, and conclusion, and appropriately including the mathematics vocabulary. In the introduction, introduce the problem and what you are being asked to do. Briefly state the steps you need to use to solve the problem. In the body, for each step, elaborate on the specific things you did, including the answer you got for each step. In the conclusion, briefly restate the type of problem and what needed to be completed. State the final answer. (e.g., "In summary, this problem required long division, for which I needed to multiply, subtract, and regroup. My final answer to the problem was 37").

In subsequent lessons, the tutors used an interactive and flexible mixture of modeling and guided practice, based on the needs of the student. If a student needed more assistance, the tutors had the option to completely model the next few lessons in the same way that Lesson 1 was modeled. If the student could do more of the work, the tutor encouraged the student to complete steps and provided reminders for each step of the sequence. We provided tutors with the latitude to revert back to modeling for any lesson in which the student needed help with the mathematics content of the lesson. When this occurred, it was often during the first exercise of the lesson, which involved the problem completed by Sam. The mathematics problem in the second exercise was always a variation of the mathematics problem in the first exercise, so the student often completed this problem with more independence.

Tutors could also determine at which lesson to begin allowing participants to complete both exercises independently. We decided this would provide a context that is most like typical classroom instruction. When moving to independent performance, the tutor monitored the student as he or she completed the steps of the problem, prompting only when absolutely necessary, or when the student directly asked for reminders. The student worked both exercises of the set independently, including all of the mathematics and writing. Again, we permitted tutors to revert to modeling or guided practice if the student was struggling with the mathematics content for a particular lesson, or if the tutor deemed the participant was moved to independent performance too soon.

Informational writing comparison condition. Students assigned to the comparison condition received expository writing instruction in science and social studies using text structures. Aside from the content and writing features taught, the writing instruction for the comparison group was similar to the writing instruction for the treatment group in several ways.

The same tutors provided instruction to the comparison students using a similar instructional approach to the MW treatment (i.e., modeling, guided practice, independent practice), students in the comparison group wrote two passages per day, and instruction was provided individually or in dyads. Comparison students also received instruction over twelve 30-minute lessons. Different from the MW intervention, we designed the informational text writing lessons to teach students to write using description, compare-contrast, sequence, cause-effect, and problem-solution text structures.

Tutor Training

RAs included six preservice teachers and one retired teacher. The six preservice teachers provided instruction as the intervention tutors. The retired teacher collected fidelity data, and substituted to provide instruction if one of the preservice teachers was absent. The first and third authors trained the RAs to teach both treatment conditions. Tutors participated in a two-hour training session in the weeks leading up to the study, and in a one-hour booster session just before the study began. The tutors were also paid for 30 min of preparation time before each lesson to review lesson materials each day. The authors randomly assigned participants from both the treatment and comparison group to each RA. The authors informed the RAs that both writing treatment conditions were expected to improve student writing, but the RAs were blind to the specific research questions. Due to distinct differences in the writing of treatment and comparison students and systematic instructional procedures, treatment diffusion was unlikely.

Data Analysis

We evaluated differences between the MW treatment and informational writing control conditions on post-test outcomes using a regression-based approach. We entered the pretest score as a control variable in the multiple regression model to account for differences at pretest.

Additionally, we included gender in the models, due to pre-intervention differences found between the treatment groups on this demographic variable. Because students from three grade levels were included in the study, we also included student grade level as a predictor.

Due to the partially nested nature of the instructional groups (some students were taught one-on-one, and some students were taught in dyads), we examined Design Effects to determine whether multi-level modeling was needed for the analyses, or whether a single-level model was warranted, using the following formula suggested by Satorra and Muthen (1995):

$$\text{Design Effect} = 1 + (\text{average cluster size} - 1) * \text{ICC}$$

When the design effect is less than 2, single level analysis of multilevel data does not generally lead to misleading results (see Maas & Hox, 2005, or Satorra & Muthen, 1995). The average cluster size for our study was only 1.37 (many students received instruction as individuals), and the design effects for each of the outcomes measures in ascending order were (a) 1.00 for the MW mathematics score—the smallest design effect possible, (b) 1.11 for %CMWS, (c) 1.17 for the MW total score, and (d) 1.24 for the MW writing score. Based on this, we decided to run single-level regression analyses.

In the models, the pretest covariates were mean-centered so that the intercept (B_0) is interpreted as the mean for the MW treatment group when the pre-test score is average. Simple regression of a continuous outcome onto a binary predictor (i.e., experimental dummy variable) is mathematically equivalent to an independent samples *t*-test. Cohen's *d* effect sizes were computed based on the unstandardized regression coefficient for condition and the standard deviation of the outcome variable. In other words, we essentially divided *B* by the pooled standard deviation of the post-test, resulting in an effect size representing the conditional effect when controlling for the covariates used in the model (see Lipsey & Wilson, 2001). Because the standardized mean difference effect size (*d*) is upwardly biased in small samples, a small sample correction was applied to the effect size, resulting in Hedges' *g* (Hedges, 1981).

Results

Descriptive statistics for the two treatment conditions are provided in Table 3. Chi-square analysis resulted in statistically significant differences between the treatment conditions on gender ($\chi^2_{(1)} = 5.11, p = .02$), indicating males and females were not distributed proportionally across the treatment groups. We identified no differences between the groups on ethnicity ($\chi^2_{(4)} = 3.02, p = .55$), free or reduced lunch status ($\chi^2_{(1)} = 0.04, p = .85$), or IEP status ($\chi^2_{(1)} = 2.59, p = .11$), indicating that these characteristics of students were distributed proportionally across the treatment conditions. We noted no statistically significant differences between the conditions on the pretest measure of writing performance ($t = -0.40, p = .692$).

We also calculated the descriptive statistics for all of the pretest and post-test scores for all outcome measures (see Table 3), and compared the treatment conditions on the pretest measures of each outcome. There were no statistically significant differences between the two conditions on pretest measures of the MW total score ($t = -1.82, p = .073$) or %CMWS ($t = 0.21, p = .829$). Pairwise correlations, with Bonferroni adjusted significance thresholds, were calculated for each of the pretest and post-test outcome measures.

Due to the small sample, we included pretest scores of the outcome variable for each of the regression models, despite the lack of a statistically significant difference between the group means. Gender was also included in the regression models, to control for potential gender differences that might have been present due to non-proportional distribution of the males and females across the conditions. Finally, grade level was included as a predictor to control for potential grade level differences in writing skill or mathematics content knowledge.

For descriptive purposes, and because we did not have a screening measure of mathematics performance, we compared the pretest scores on the mathematics-writing measure to scores of students in prior studies. The treatment and control groups in the current study received mean total scores of 1.21 (*SD* = 0.25) and 2.27 (*SD* = 0.54), respectively. In a previous study by Powell & Hebert (2016), students at the 25th percentile or below on the Wide Range Achievement Test-4 math computation subtest scored received mean scores of 1.6 on (*SD* = 2.07) on the same math-writing measure, whereas students above the 25th percentile received mean scores of 5.89 (*SD* = 3.76). In another study by Hebert, Powell, Bohaty and Roehling (under review), students in grades 3 to 5 with a grade equivalent score of one or more grades below their current grade level on the Woodcock Johnson-III (WJ-III) Mathematics Fluency subtest received mean scores of 1.09 (*SD* = 1.81), whereas students who had grade equivalent scores on the WJ-III that were more than a grade level below their current grade level had scores of 3.32 (*SD* = 3.01). These comparisons provide further evidence that the participants fit the criteria of the target population for the current study.

Mathematics-Writing Scores

As previously described, raters scored students' responses to the MW measures for *writing organization* and *mathematics content*, and then combined those two scores to obtain a *total* MW score. We first examined the effect of the intervention on the overall MW score, and followed this up by examining whether there were impacts on writing organization, and mathematics content, separately (see Table 4).

Total score. The regression analyses revealed a statistically significant effect of treatment on the MW total score. Students in the MW intervention scored, on average, 2.90 points higher than students in the comparison condition. Gender and grade level were not significant predictors of the total score, indicating these variables did not predict the total score outcome. The pretest, however, was a significant predictor of the outcome in the model, indicating that an increase of 1 point on the pretest measure predicted a corresponding increase of 0.54 points on the MW total score. Thus, the resulting standardized mean difference between groups of $d = 1.05$ [95% *CI* = 0.51, 1.58] was conditional on the pretest covariates, and was calculated using the standardized beta-weight and the standard deviation of the dependent variable.

Writing organization score. The regression analyses also revealed a statistically significant effect of treatment on the writing organization score of the MW measure. Students in the MW condition scored, on average, 2.94 points higher than students in the control condition on the writing organization score. The resulting standardized mean difference between groups of $d = 1.49$ [95% *CI* = 0.92, 2.06] was conditional on the pretest covariates, and was calculated using the standardized beta-weight and the pooled standard deviation of the dependent variable. Gender was not a significant predictor of the organization score. Grade level and pretest were both significant predictors in the model. This indicated that students scored 0.62 points higher for each increase in grade level, and that an increase of 1 point on the pretest measure predicted a corresponding increase of 0.62 points on the writing organization score.

Mathematics content score. The regression analyses indicated there was no statistically significant effect of treatment on the mathematics content score. Students in the MW experimental condition scored 0.14 points higher, on average, than students in the comparison condition on the mathematics outcome, but the resulting standardized mean effect size for treatment of $d = 0.11$ [95% *CI* = -0.40, 0.61] was not large enough to be detected statistically. Grade level and gender were not significant predictors of the mathematics content score, although grade level would have been significant if the *p*-value threshold for significance was set at .10, indicating that it may be worthwhile to follow up on this variable in future studies with more power. Pretest was also a significant predictor in the model, indicating that an increase of 1 point on the mathematics content pretest measure predicted a corresponding increase of 0.30 points on the mathematics content score.

%CMWS

The results of the analyses indicated statistically significant differences on %CMWS between the MW intervention and informational writing comparison conditions. The %CMWS was 20% higher for the MW condition than in the comparison after controlling for pretest scores, gender, and grade level. The standardized beta-weight was used to calculate the effect size of treatment, and was also conditional on the pretest covariates. The resulting standardized mean difference between treatment groups was *d* = 0.82 [95% *CI* = 0.30, 1.34]. Gender and grade level were not statistically significant predictors of the %CMWS measure. The pretest, however, was a significant predictor of the outcome, indicating that every increase of 10%CMWS on the pretest predicted an increase of 2.8%CMWS on the post-test. Table 5 includes the results of the regression model for %CWMS.

Discussion

We designed this study to explore the effects of a MW intervention on the MW of students completing grades 3, 4, and 5, who were also at-risk for a learning disability. It was expected that students with low-writing skills may also have difficulty writing about mathematics, due to moderate correlations between writing scores and MW scores in previous studies (Hebert & Powell, 2016; Powell & Hebert, 2016). There were no significant correlations between pretest writing scores and pretest scores of MW. The MW intervention led to improvements in MW scores and CBMs, indicating there may be relationships among these variables.

Writing Organization and Mathematics Content

The intervention was designed with a focus on the writing organization aspects of MW. Mathematics content was included in the instruction, but varied widely across lessons and was not explicitly taught. This is reflected in the results, which indicated the intervention led to improvements in MW organization but not in mathematics content. The increase in %CMWS indicated students may have written about mathematics using better conventions, but that does not show an increase in mathematics knowledge.

Our results are encouraging, in that clearer organization of writing is likely to ensure that students' knowledge is accurately represented within a writing sample. As such, teachers may be more likely to be able to discern whether students addressed each component of the problem, and teachers may have a stronger sense of where misunderstandings in mathematical concepts or procedures occurred. Despite those potential advantages, clearer organization did not result in higher mathematical content scores in the sample for the current study. This indicates that, although the students may have had more organized writing, it did not necessarily show areas in which students had knowledge they were not able to express during the pretest.

These results are not necessarily surprising, as the mathematics content varied across each lesson (see Table 2) and was not directly related to the mathematics content on the outcome measure. Due to this variation, the intervention was not likely to increase mathematical content knowledge (conceptual or procedural) in a specific area that could be applied to when writing about mathematics. Thus, we would not expect increases in mathematics knowledge.

Additionally, it may be that MW organization also has both general and specific qualities.

That is, there may be ways to improve general organization when responding to any MW prompt, as well as ways to improve the organization of writing for specific mathematics content. For example, this study focused on teaching students to write about mathematics by generally organizing writing to (a) identify what the problem is asking, (b) explain mistakes of the pseudo-student, (c) explain how to help the pseudo-student fix the mistakes, and (d) conclude with a short summary of how they helped the student. That type of general framework is helpful for the writer and reader, especially when writing about the mistakes made by a pseudo-student. A general writing framework, however, does not address specific mathematics content that may benefit from more specific organization, such as (a) how to organize a discussion for explaining the process used for converting numerators and denominators when writing about fractions, (b) describing conceptual rationales before explaining procedures in algebra (e.g., “I need to isolate the unknown variable to solve for it. To do that…”), or (c) describing a geometrical formula before discussing how it can be used to solve a problem. More work is needed to determine whether pairing our instructional approach with specific mathematics content may also impact mathematics knowledge and lead to greater gains.

Percentage of Correct MW Sequences (%CMWS)

In addition to improvements in overall organization, this study also led to improvements in the CBM, a measure of writing conventions for the use of numbers, symbols, and equations in MW along with spelling, grammar, syntax, and punctuation. The %CMWS score was used to control for both the length of student writing and the time used to complete the writing sample. Therefore, the positive effects for this measure indicate that students were better able to incorporate numbers and symbols into their writing with fewer syntactical errors. While specific mathematics was not taught, the use of numbers, symbols, and equations within written text was modeled and supported throughout the intervention. Therefore, it was encouraging to see improvements. Our result demonstrates that a brief intervention can lead to improvements in students' ability to communicate in writing with numbers, symbols, and equations, which is important to demonstrating knowledge about mathematics.

Like the discussion of improvements in writing organization, it is more likely that students will be able to express their knowledge about mathematics if they can use writing conventions appropriately when combining words and mathematical terms to write about text. This alone, however, did not lead to better scores on the mathematics content measure, indicating that combining this instruction with specific mathematics content may be necessary.

The pilot of the %CMWS measure was also an important outcome for this study. This may be one of the first attempts at developing a CBM measure for MW, and it is encouraging that the research assistants could be trained to score with high inter-rater reliability. The scores also correlated moderately with our other measure of MW (r = .49), providing some early validity evidence for this measure. Future research needs to expand on the use of CBMs for MW, with conventions determined through further piloting and theoretical development.

Some decisions we made could be considered controversial. For example, we elected to score incorrect word sequences when students combined numbers and word within the same equation (e.g., "two plus 2 = 4"). We also decided that symbols, such as dollar signs and numbers, would be considered as separate word sequences when paired, rather than counting them as a single unit. For example: *^The^wand^costs^$^4^.* (6 word sequences) rather than, *^The^wand^costs^$4^.* (5 word sequences). Such decisions may be debated in the mathematics and writing communities, and consensus must be reached about which conventions are appropriate for scoring. Regardless of the scoring decisions, the statistically significant results are not likely to be an artifact of the scoring, as the same rules were applied to the writing of both the MW intervention and comparison conditions at both time points, and pretests and post-tests were mixed during scoring to avoid bias.

Implications for Practice

The research findings in this paper indicate that MW instruction can improve the performance of students with learning difficulties on assessments requiring them to write about mathematics. The ES for mathematics content was not statistically significant; it favored the treatment group, suggesting students receiving MW instruction may improve their ability to demonstrate their mathematics knowledge on such assessments. Although more research is needed, pairing MW instruction with mathematics content instruction could improve the ability of students with learning difficulties to communicate their mathematics knowledge. Teachers should consider including MW instruction for students with learning difficulties into their classrooms.

Limitations and Future Research

In this study, we compared the MW intervention to an informational text intervention. While this arrangement allowed us to control for minutes spent with a tutor, it could be considered a weakness of the study. Future research should compare the performance of students in the MW intervention to students in a mathematics-content intervention, or to a business-as-usual comparison (i.e., no intervention).

Another limitation of the study was that the mathematics content varied for each lesson in the intervention rather than focusing on specific mathematics content. This led to an important finding - that MW content scores may not necessarily improve simply by engaging in MW. Future research should explore whether MW within a specific content area leads to positive mathematics content outcomes. With only twelve 30-minute lessons, this intervention was relatively short in duration. Therefore, the full potential of such an intervention to impact MW organization and content knowledge may not have been fully realized. It is possible that additional instruction may have led to larger impacts on organization and conventions, or to statistically significant impacts on mathematics content knowledge. Therefore, further exploration of MW instruction should examine the effects of longer interventions. Future research should also investigate whether a mathematics-writing intervention needs to be coupled with a mathematics-content intervention (i.e., instruction focused on improving mathematics conceptual and procedural knowledge) to boost student outcomes in both mathematics and mathematics writing. This research should also include measures of mathematics content knowledge. As we stated previously, we did not find impacts on the math context score for the MW content measure, and we stated there is not a theoretical reason to expect an impact from writing instruction alone. However, this is an empirical question that should be examined further.

The small sample of this pilot study was also a limitation. Although the effect size for mathematics content knowledge was small and was not statistically significant, it is important to follow this up with a study with a larger sample in order to determine whether there are some impacts of better writing organization and conventions on students' ability to express their content knowledge. Such follow-up with a larger sample would be critical to provide evidence-based mathematics-writing strategies for teachers and researchers. It is clear that future MW interventions need to focus on specific content to improve students' content knowledge, but it is also important to understand whether there are unique contributions of better organization and conventions to students' ability to communicate about mathematics through writing. Such knowledge is important for developing the most impactful intervention possible. For example, if there are no impacts of writing organization or conventions on students' ability to express their content knowledge, it may not be necessary to address these skills in an intervention. A larger study is needed to determine this.

Differential attrition was also a limitation of this study. Based on the information provided by families for having to drop out of the study prior to the start (e.g., summer vacation schedule conflicts), we believe the attrition to be unrelated to the study. It is possible, however, the attrition was systematic in some unidentifiable way that may have impacted the results of the study. Because this occurred prior to the study, we were not able to compare the assessment scores of participants who completed the study to scores for those who were dropped from the study to determine whether there were important differences between the groups.

Finally, we did not determine whether participation in the mathematics-writing intervention led to improved high-stakes testing outcomes for students. Future research should determine the level of transfer from intervention to high-stakes performance, and should examine the degree to which a mathematics-writing intervention can influence mathematics-writing responses on a high-stakes test.

Conclusion

This pilot study resulted in important findings illustrating the effectiveness of teaching MW skills on writing organization and conventions. The statistically significant effect sizes were moderate to large, a finding that is noteworthy considering the short duration of the intervention. A first attempt to examine MW conventions through the development of %CMWS as a CBM is also promising. Exploration of MW assessments and interventions is still a relatively new area of research, and much work needs to be done to develop theory and investigate the mechanisms underlying the development of these skills. This study provides an important step in the iterative development of theories, assessments, and interventions in the area of MW.

References

Andersson, U. (2010). Skill development in different components of arithmetic and basic cognitive functions: Findings from a 3-year longitudinal study of children with different types of learning difficulties. *Journal of Educational Psychology, 102*, 115-134. https://doi.org/10.1037/a0016838

Barlow, A. T., & Drake, J. M. (2008). Division by a fraction: Assessing understanding through problem writing. *Mathematics Teaching in the Middle School, 13*, 326-332.

Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Writing problems in developmental dyslexia: Under-recognized and under-treated. *Journal of School Psychology, 46*, 1-21. https://doi.org/10.1016/j.jsp.2006.11.008

Breaux, K. C. (2010). *Wechsler individual achievement test—third edition: Technical manual*. Bloomington, MN: Pearson.

Cirino, P. T., Fuchs, L. S., Elias, J. T., Powell, S. R., & Schumacher, R. F. (2015). Cognitive and mathematical profiles for different forms of learning difficulties. *Journal of Learning Disabilities, 48, 156–175*. https://doi.org/10.1177/0022219413494239

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

Evens, H., & Houssart, J. (2004). Categorizing pupils' written answers to a mathematics test question: 'I know but I can't explain.' *Educational Research, 46*, 269-292. https://doi.org/10.1080/0013188042000277331

Fletcher, J. M. (2005). Predicting math outcomes: Reading predictors and comorbidity. *Journal of Learning Disabilities, 38*, 308-312. https://doi.org/10.1177/00222194050380040501

Fuchs, L. S. & Fuchs, D. (2002). Mathematical problem-solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities, 35*, 564-574. https://doi.org/10.1177/00222194020350060701

Gillespie, A., & Graham, S. (2014). A meta-analysis of writing interventions for students with learning disabilities. *Exceptional Children*, *80*, 454-473. https://doi.org/10.1177/0014402914527238

Hebert, M., Kearns, D. M., Hayes, J. B., Bazis, P., & Cooper, S. (2018). Why Children with Dyslexia Struggle with Writing and How to Help them. *Language, Speech, and Hearing Services in Schools, 49*, 843-863. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0024

Hebert, M. & Powell, S. (2016). Examining fourth grade mathematics writing: Features of Organization, Mathematics Vocabulary, and Mathematical Representations. *Reading and Writing: An Interdisciplinary Journal, 29*, 1511-1537. https://doi.org/10.1007/s11145-016-9649-5

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107-128. https://doi.org/10.3102/10769986006002107

Hughes, C. A., Morris, J. R., Therrien, W. J., & Benson, S. K. (2017). Explicit instruction: Historical and contemporary contexts. *Learning Disabilities Research and Practice, 32*, 140-148. https://doi.org/10.1111/ldrp.12142

Jigyel, K., & Afamasaga-Fuata'i, K. (2007). Students' conceptions of models of fractions and equivalence. *Australian Mathematics Teacher, 63*(4), 17-25.

Kasmer, L. A., & Kim, O.-K. (2012). The nature of student predictions and learning opportunities in middle school algebra. *Educational Studies in Mathematics, 79*, 175-191. https://doi.org/10.1007/s10649-011-9336-z

Kostos, K., & Shin, E.-K. (2010). Using math journals to enhance second graders' communication of mathematical thinking. *Early Childhood Education Journal, 38*, 223-231. https://doi.org/10.1007/s10643-010-0390-4

Koutsoftas, A. D., & Gray, S. (2012). Comparison of narrative and expository writing in students with and without language-learning disabilities. *Language, Speech, and Hearing Services in Schools, 43*, 395-409. https://doi.org/10.1044/0161-1461(2012/11-0018)

Landerl, K., Fussenegger, B., Moll, K., & Willburger, E. (2009). Dyslexia and dyscalculia: Two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology, 103*, 309-324. https://doi.org/j.jecp.2009.03.006

Lichtenberger, E. O., & Breaux, K. C. (2010). *Essentials of psychological assessment series: Essentials of WIAT-III and KTEA-II assessment.* Hoboken, NJ: John Wiley & Sons.

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

McMaster, K. L., Du, A., Parker, D. C., & Pinto, V. (2011). Using curriculum-based measurement for struggling beginning writers. *Teaching Exceptional Children, 44*(2), 26-34. https://doi.org/10.1177/004005991104400203

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in writing. *Journal of Special Education, 41*, 68-84. https://doi.org/10.1177/00224669070410020301

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities, 40*, 458-478.

https://doi.org/10.1177/00222194070400050901

Satorra, A., & Muthen, B. (1995). Complex sample data in structural equation modeling. *Sociological methodology, 25*, 267-316.

National Center for Education Statistics. (2011). *National Assessment of Educational Progress, 2011 Writing Assessment*. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Dept. of Education.

National Governors Association Center for Best Practices & Council of Chief State School Officers (2010). *Common Core State Standards mathematics.* Washington, D.C.: Authors.

Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing, 26*, 45-65. https://doi.org/10.1007/s11145-012-9392-5

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new U.S. intended curriculum. *Educational Research, 40*, 103-116. https://doi.org/10.3102/0013189X11405038

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

Powell, S. R., Hebert, M. A., Cohen, J. A., Casa, T. M., & Firmender, J. M. (2017). A Synthesis of Mathematics Writing: Assessments, Interventions, & Surveys. *Journal of Writing Research, 8*, 493-526. https://doi.org/10.17239/jowr-2017.08.03.04

Powell-Smith, K. A., & Shinn, M. R. (2004). Administration and scoring of Written Expression Curriculum-Based Measurement (WE-CBM) for use in general outcome measurement:

AIMSWeb training workbook. Eden Prairie, MN: Edformation. Retrieved from: https://studylib.net/doc/8259685/administration-and-scoring-of-written-expression

Psychological Corporation (2009). *Wechsler Individual Achievement Test (3rd edition).* San Antonio, TX: Author.

Ritchey, K. D., & Coker, D. L. (2014). Identifying writing difficulties in first grade: An investigation of writing and reading measures. *Learning Disabilities Research and Practice, 29*, 54-65. https://doi.org/10.1111/ldrp.12030

Swanson, H. L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 100*, 343-379. https://doi.org/10.1037/0022-0663.100.2.343

Troia, G. A. (2006). Writing instruction for students with learning disabilities. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.) *Handbook of writing research* (pp. 324-336). New York: Guilford.

Vukovic, R. K., & Siegel, L. S. (2010). Academic and cognitive characteristics of persistent mathematics difficulty from first through fourth grade. *Learning Disabilities Research and Practice, 25*, 25-38. https://doi.org/10.1111/j1540-5826.2009.00298.x

Table 1

Demographic Characteristics by Group

	Mathematics-Writing (treatment)	Information Writing (control)
N	29	32
Grade		
3	7	9
4	11	11
5	11	12
Percent Female	28%	56%
Ethnicity		
Asian	10%	3%
Black/African American	4%	13%
Caucasian	72%	69%
Hispanic/Latinx	10%	9%
Two or more races	4%	6%
FRL	21%	19%
IEP	24%	44%
WRMT-III (Basic Skills)	83.44 (12.72)	89.00 (13.75)
WRMT-III (Reading Comprehension)	88.40 (15.52)	92.38 (13.92)
WIAT-III Essay Composition (Pretest)	96.38 (13.13)	94.91 (15.51)

Note. FRL = Free or Reduced Lunch; IEP = Individualized Education Plan; WRMT-III = Woodcock Reading Mastery Test, 3rd edition; WIAT-III = Wechsler Individualized Achievement Test, 3rd edition.

Table 2

Outline of Mathematics-Writing Intervention

Lesson	Mathematics content	Domain area	Difficulty level
1	Solve comparison and total word problem	Algebra	Hard
2	Use place value to determine value	Number properties and operations	Hard
3	List events in order they occurred	Algebra	Easy
4	Identify missing information to solve a problem	Number properties and operations	Hard
5	Determine perimeter of a rectangle	Measurement	Easy
6	Compare unit fractions in context	Number properties and operations	Medium
7	Multiply 5-digit number by 2-digit number in context	Number properties and operations	Medium
8	Reason about relationships to reach conclusions	Algebra	Easy
9	Determine which rectangular floor has greatest area	Measurement	Hard
10	Solve a story problem involving division	Number properties and operations	Easy
11	Solve a story problem involving multiplication	Number properties and operations	Medium
12	Solve a story problem involving time	Measurement	Medium

Table 3

Outcome Measures by Condition

	Mathematics-Writing (*n* = 29)	Informational text comparison (*n* = 32)
Measures	*M* (SD)	*M* (SD)
MW Measure		
Mathematics score (pretest)	1.48 (2.10)	0.78 (1.01)
Mathematics score (post-test)	1.50 (1.43)	1.16 (1.32)
Writing score (pretest)	0.79 (1.26)	0.44 (0.76)
Writing score (post-test)	4.11 (2.35)	0.81 (1.35)
Total score (pretest)	2.28 (2.93)	1.21 (1.41)
Total score (post-test)	5.61 (2.87)	1.97 (2.36)
%CMWS		
Pretest	0.56 (0.28)	0.58 (0.37)
Post-test	0.74 (0.14)	0.53 (0.32)
WIAT-III Essay Composition (posttest)	98.00 (13.16)	96.59 (15.07)

Note. WIAT-III = Wechsler Individualized Achievement Test, 3rd Edition

Table 4

Regression Results for Mathematics-Writing Measures

Model/Parameter	Unstandard coefficient (B)	SE	*t*	*p*-value
Mathematics score ($R^2 = .15$)				
Intercept	1.14	0.93	1.23	.225
Treatment	0.14	0.37	0.39	.698
Gender	0.09	0.37	0.26	.799
Grade	0.01	0.22	0.06	.951
Pretest (centered)	0.30	0.11	2.72	.009
Writing score ($R^2 = .53$)				
Intercept	-1.37	1.23	-1.11	.271
Treatment	2.93	0.49	6.05	<.001
Gender	-0.47	0.48	-0.97	.337
Grade	0.62	0.29	2.13	.037
Pretest (centered)	0.62	0.22	2.75	.008
Total Score ($R^2 = .50$)				
Intercept	0.22	1.64	0.13	.895
Treatment	2.90	0.66	4.42	<.001
Gender	-0.51	0.65	-0.78	.436
Grade	0.56	0.38	1.47	.147
Pretest (centered)	0.54	0.14	3.90	<.001

Note. For condition, comparison (informational text) = 0 and MW = 1.

Table 5

Regression Results for %CMWS

Model/Parameter	Unstandard coefficient (B)	SE	*t*	*p*-value
Post-test %CMWS (R^2 = .30)				
Intercept	0.35	0.16	2.13	.038
Treatment	0.20	0.63	3.19	.002
Gender	-0.49	0.06	-0.76	.450
Grade	0.05	0.04	1.32	.193
Pretest (centered)	0.29	0.10	3.01	.004

Note. For condition, comparison (informational text) = 0 and MW = 1; CMWS = Correct MW Sequences.

Appendix A

Rules for Scoring %CMWS

Score all writing with CWS rules. When students include numbers and symbols in text, follow these additional rules for scoring mathematics correct MW sequences. When finished, sum all correct MW sequences and incorrect MW sequences. Divide the total correct MW sequences by the total number of correct plus incorrect MW sequences.

Rule: With an equation, expression, or amount of money written using <u>words</u>, score using CWS rules.

Rule: With money written using <u>words</u>, students must include the words, “dollar(s)” and “cent(s)”.

Rule: With money written using <u>digits</u>, students should not leave a space between the dollar sign ($) and subsequent digit. Count the dollar sign and subsequent digit as one unit.

Rule: Score equations as mathematically correct if the number or expression on the right side of the equal sign (=) is equivalent to the number or expression on the left side of the equal sign.

Rule: With writing about an equation, expression, or amount of money, count the use of words such as *did*, *put*, *doing*, *took*, etc., as correct:

I *did* 8 × 5 I *put* 9 + 4 = 13
I *was doing* $20 - $12.49 I *took* $20 – $12.49

Rules for scoring equations:

1. Students may write the equation, expression, or money amount in either numbers and symbols or words. They must, however, use the choice consistently through the entire equation, expression, or money amount:

Examples
1 + 1 = 2 one plus one equals two
Non-examples
one plus 1 = 2 1 plus one equals two one plus one equals 2
$12.49 12 dollars and 49 cents
twelve forty-nine plus $7 12 dollars and 49¢

2. If the equation, expression, or money amount alternates between numbers and symbols and words, the first unit following a switch is an error:

one plus 1 = 2 1 plus one equals two one plus one equals 2
twelve dollars and 49¢ 12 dollars and 49 cents
12 dollars and forty-nine cents twelve forty-nine plus $7