Alternative Paths to Improved Word-Problem Performance: An Advantage for Embedding Pre-Algebraic Reasoning Instruction Within Word-Problem Intervention

Sarah R. Powell[1], Katherine A. Berry[1], Anna-Maria Fall[1],

Greg Roberts[1], Lynn S. Fuchs[2], and Marcia A. Barnes[2]

[1]The University of Texas at Austin

[2]Vanderbilt University

Powell, S. R., Berry, K. A., Fall, A.-M., Roberts, G., Fuchs, L. S., & Barnes, M. A. (2021). Alternative paths to improved word-problem performance: An advantage for embedding pre-algebraic reasoning instruction within word-problem intervention. *Journal of Educational Psychology, 113*(5), 898–910. https://doi.org/10.1037/edu0000513

Author Note

This research was supported in part by Grant R324A150078 from the Institute of Education Sciences in the U.S. Department of Education to the University of Texas at Austin. The content is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. Department of Education.

Correspondence concerning this article should be addressed to Sarah R. Powell, 1 University Station D5300, Austin, TX 78712. E-mail: srpowell@austin.utexas.edu

Abstract

The purpose of this study was to explore the paths by which word-problem intervention, with versus without embedded pre-algebraic reasoning instruction, improves word-problem performance. Students with mathematics difficulty (MD; $n = 304$) were randomly assigned to a business-as-usual condition or one of two variants of word-problem intervention. The pre-algebraic reasoning component targeted relational understanding of the equal sign as well as standard and nonstandard equation solving. Intervention occurred for 16 weeks, 3 times per week, 30 min per session. Sequential mediation models revealed main effects, in which each intervention condition significantly and substantially outperformed the business-as-usual condition, corroborating prior research on the efficacy of schema word-problem intervention. Yet despite comparable effects on word-problem outcomes between the two word-problem conditions, the process by which effects accrued differed: An indirect path via equal-sign understanding and then equation solving was significant only for the word-problem intervention condition with embedded pre-algebraic reasoning instruction. Additionally, the effect of this condition on equal-sign reasoning was strong. Given the link between equal-sign reasoning for success with algebra and the importance of algebra for success with advanced mathematics, results suggest an advantage for embedding pre-algebraic reasoning instruction within word-problem intervention.

Key words: learning difficulty; mathematics; pre-algebra; word problems

Educational Impact and Implications Statement

This study suggests pre-algebraic reasoning is important within math word-problem instruction for 3rd-grade students who experience difficulty with math. Pre-algebraic reasoning involves interpreting the equal sign as “the same as” and solving equations (e.g., 3 + __ = 9 or 7 = 13 – __). As students develop strong pre-algebraic reasoning, they are better equipped to solve word-problems.

Alternative Paths to Improved Word-Problem Performance: An Advantage for Embedding Pre-Algebraic Reasoning Instruction Within Word-Problem Intervention

To demonstrate mathematics competency in the elementary grades, students must interpret and solve word problems, and many students with mathematics difficulty (MD) experience less success solving word problems than students without MD (Swanson et al., 2014). Importantly, students' competence in solving word problems predicts success in algebra, a phenomenon documented in individual difference studies (Powell & Fuchs, 2014) and reported in a large national survey of United States (U.S.) Algebra I teachers (Hoffer et al., 2007). Further, a large-scale randomized controlled trial demonstrated that word-problem intervention improved pre-algebra performance more than calculations intervention (Fuchs, Powell, et al., 2014; Powell, Fuchs, et al., 2015). This research suggests a link between word problems and algebra; generating and solving equations to represent word problems are transparently connected to pre-algebraic and algebraic thinking.

Even so, research is needed to explore the paths by which word-problem intervention, with versus without embedded pre-algebraic reasoning instruction, improves word-problem performance. In prior research, authors have not explicitly provided pre-algebraic reasoning instruction within a word-problem intervention (Fuchs, Powell, et al., 2014) or researchers used a pre-algebraic reasoning component limited in scope within a quasi-experiment (Powell & Fuchs, 2010). A randomized controlled trial contrasting the two variations of word-problem intervention may provide causal evidence on the role of pre-algebraic thinking within word-problem solving. Moreover, examining the path by which each variation of word-problem intervention improves word-problem solving may deepen understanding about connections between word-problem performance and pre-algebraic reasoning, while guiding word-problem intervention practices.

The purpose of the present study was to explore the paths by which word-problem intervention, with versus without embedded pre-algebraic reasoning instruction, improves word-problem performance of students with *mathematics difficulty* (MD). We defined MD as scoring at or below the 25th percentile on a word-problem screener. We used the umbrella term MD to include students with a school-identified specific learning disability and Individualized Education Program (IEP) goals in mathematics, students identified with dyscalculia, or those with persistent and below-grade level mathematics performance, as defined by other researchers (Branum-Martin et al., 2012; Bryant et al., 2016; Fuchs, Schumacher, et al., 2014). In the early elementary grades, these students experience difficulty with tasks involving counting, comparing, adding, and subtracting as compared to students without MD (Aunio et al., 2015; Geary et al., 2012; Martin et al., 2013). Students with MD also experience difficulty solving addition and subtraction word problems (Tolar et al., 2016).

Pre-Algebraic Reasoning and Students with MD

Pre-algebraic reasoning requires foundational knowledge related to counting, adding, and subtracting, as well as higher-level reasoning and interpretation. As described by Pillay et al. (1998), pre-algebraic reasoning involves understanding the equal sign as a relational symbol and solving equations with a single variable (i.e., the unknown). We relied on this definition from Pillay et al. to guide this research study.

In the U.S., mathematics standards at first grade require students to understand the relational meaning of the equal sign and solve equations with an unknown in different positions of an equation (e.g., 5 = __ – 3; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). As students progress through the elementary grades, they are expected to solve equations with an unknown across all four operations, with continuing emphasis in the standards on understanding the equal sign and solving equations with an unknown. Thus, pre-algebraic reasoning and arithmetic skill are expected to develop concurrently.

The focus on the equal sign in U.S. standards is necessary because many elementary students view the equal sign as operational; that is, defining the equal sign as the answer to a problem, the total, or to do something (Matthews et al., 2012; Stephens et al., 2013; Vincent et al., 2015). This operational understanding of the equal sign may be explicitly taught or implicitly learned as students solve a very high percentage of equations (e.g., 11 – 5 = __) that require little interpretation of the equal sign as relational (Powell, 2012). In fact, students should interpret the equal sign as a balance between two sides of an equation (Jacobs et al., 2007). Such relational understanding of the equal sign can improve with explicit instruction about the meaning of the equal sign (Matthews & Rittle-Johnson, 2009; McNeil & Alibali, 2005) or with practice solving standard and nonstandard equations with an unknown in various positions (Blanton et al., 2015; Powell, Driver, et al., 2015).

Without instruction, students with MD often provide operational definitions of the equal sign. According to Powell and Fuchs (2010), more than 90% of third-grade students with MD explained the equal sign operationally with examples such as, “what the sum is” or “an answer” (p. 391). After receiving explicit instruction on the equal sign as a relational symbol, definitions improved significantly (e.g., “the same as”). Notably, students with MD outperformed students without MD (who did not receive explicit instruction about a relational meaning of the equal sign) on equal-sign definitions and tasks (Powell & Fuchs, 2010).

A relational interpretation of the equal sign may influence how students solve equations with an unknown (Knuth et al., 2006; Matthews et al., 2012; Rittle-Johnson & Alibali, 1999). In the elementary grades, students may solve standard or nonstandard equations (Powell, 2012). Standard equations feature an operator symbol on the left side of the equal sign and the equal sign in a standard position (2 + __ = 7). Nonstandard equations feature an operator symbol on the right side of the equal sign (10 = 14 – __), on both sides of the equal sign (8 + 3 = __ + 5), or on neither side of the equal sign (4 = __). Across the elementary grades and achievement histories, students demonstrate greater success solving standard over nonstandard equations (Powell et al., 2016).

Students with MD exhibit greater difficulty solving equations than students without MD. For example, Driver and Powell (2015) administered a test of nonstandard equation solving to second graders. Students with MD scored significantly lower than students without MD (ES [effect size] = 1.01). With explicit instruction on solving standard and nonstandard equations, students with MD showed improvement (for standard equations, ES = 0.78; for nonstandard equation solving, ES = 1.54; Powell, Driver, et al., 2015). Importantly, practice with solving both standard and nonstandard equations also contributed to improved understanding of the equal sign as a relational symbol (ES = 1.05).

Pre-algebraic reasoning serves as a prerequisite for understanding algebra (Pillay et al., 1998), and algebra acts as a gatekeeper to later success in school and career (Byun et al., 2015; Eddy et al., 2015). In the elementary grades, pre-algebraic reasoning skills include the ability to interpret the equal sign as relational and success with solving standard and nonstandard equations. Typically, understanding the equal sign as relational is foundational for balancing two sides of an equation (Knuth et al., 2006; Matthews et al., 2012; Rittle-Johnson & Alibali, 1999). However, exposure to and practice with nonstandard equations can improve relational understanding of the equal sign (McNeil et al., 2011; Powell, Driver, et al., 2015). Students with MD, who demonstrate lower pre-algebraic performance than students without MD, may require supplemental pre-algebraic instruction. This pre-algebraic instruction may prove important within word-problem solving because students generate and solve standard and nonstandard equations to set up and solve word problems (Xin et al., 2011).

Word-Problem Solving and Students with MD

A word problem is a text-based mathematics problem in which students use information from the problem to answer a question about a missing quantity (e.g., “Suzy ran for 23 minutes during the soccer game. Xin ran 16 minutes more than Suzy. How many minutes did Xin run during the game?”). In the U.S., expectations for students to solve word problems appear in mathematics standards as early as kindergarten (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). By third grade, students are expected to set up and solve word problems during classroom instruction and on high-stakes tests.

Although word-problem solving is demanding for many students (Fuchs, Powell, et al., 2014), students with MD often demonstrate markedly poorer performance relative to students without MD (Peake et al., 2015; van Garderen et al., 2012). Fortunately, word-problem outcomes for students with MD can improve with word-problem intervention (Flores et al., 2016; Fuchs, Powell, et al., 2014; Griffin et al., 2018; Swanson et al., 2014; Xin et al., 2011; Zheng et al., 2012).

Word-problem intervention focused on schemas has proven successful for students with MD (Cook et al., 2019; Fuchs, Powell, et al., 2014; Jitendra et al., 2015; Peltier & Vannest, 2017). A *schema* refers to the word-problem structure or word-problem type (e.g., this problem is about parts and a total; that problem compares two amounts for a difference). In the early elementary grades, students solve word problems featuring three additive schemas: Total, Difference, and Change (García et al., 2006; Griffin & Jitendra, 2009; Kintsch & Greeno, 1985; Willis & Fuson, 1988). At second grade, Powell, Fuchs, et al. (2015) provided explicit schema instruction to students with MD. Students who received the word-problem intervention focused on additive schemas demonstrated significant gains on word-problem measures over students in an intervention focused on calculation skill and over the business-as-usual (BaU) comparison. Fuchs, Seethaler, et al. (2008) found similar results with a sample of third-grade students with MD. Researchers also have identified positive effects of multiplicative schema instruction in late elementary school and middle school for students with MD (Jitendra et al., 2017; Xin & Zhang, 2009).

Connection Between Pre-Algebraic Reasoning and Word-Problem Solving

Most central to the present study, Powell and Fuchs (2010) conducted a quasi-experiment in which 80 students with MD received word-problem intervention with pre-algebraic reasoning, word-problem intervention without pre-algebraic reasoning, or BaU. Students in the two intervention conditions received 15 sessions focused on the Total schema. In the condition with pre-algebraic reasoning instruction, 2 to 4 min of each session focused on interpreting the equal sign as a relational symbol and identifying (but not solving) correct and incorrect closed equations (e.g., $3 = 7 – 4$ is correct whereas $4 = 9 + 5$ is incorrect). Interventionists taught students to analyze equations and determine if one side of an equation was the same as the other side. To control for intervention time, students in the word-problem intervention condition without pre-algebraic reasoning instruction received 2 to 4 min of extra practice solving calculation problems. A mediation analysis revealed that the word-problem intervention with pre-algebraic reasoning led to improved performance on a measure of nonstandard equations, which in turn was associated with improved word-problem outcomes.

Purpose and Research Questions

The purpose of the present study was to explore the paths by which word-problem intervention, with versus without embedded pre-algebraic reasoning instruction, improves word-problem performance. This study extends Powell and Fuchs's (2010) quasi-experiment in a variety of ways. The present study expanded and reconfigured the pre-algebraic reasoning instruction component, newly providing practice with a balance scale and manipulatives to deepen students' understanding of the equal sign as relational; encouraging students to generate drawings to solve equations; and modeling and practice in isolating the unknown variable on one side of the equal sign. Also, word-problem intervention, into which we embedded pre-algebraic reasoning instruction, included a focus on all three additive word-problem schemas, instead of the solitary focus on Total problems in Powell and Fuchs (2010, as well as a longer intervention duration.

Methodologically, the present study extends Powell and Fuchs (2010) in two critical ways, designed to deepen insight into connections between pre-algebraic reasoning and word-problem performance and to provide direction for practice. First, the present study was a randomized controlled trial, rather than a quasi-experiment. Second, this study adopted an expanded focus on mediation. We considered equal-sign understanding and equation solving with standard and nonstandard equations as sequential mediators (versus the sole use of nonstandard equation solving as the only mediator in Powell and Fuchs), in which effects of the embedded component were expected to improve equal-sign understanding. Understanding the equal sign as relational is foundational for balancing and solving equations (Matthews et al., 2012; Rittle-Johnson & Alibali, 1999). For example, Byrd et al. (2015) determined a relational interpretation of the equal sign was essential for solving equations correctly. With this information, we hypothesized sequential mediation from equal-sign knowledge (*equal sign*) to equation solving (*open equations*). As equal-sign understanding improved, we expected equation-solving improvement; and which we expected, in turn, to be associated with stronger word-problem outcomes. Moreover, we tested mediation not only for the contrast between the word-problem intervention condition with the pre-algebraic component versus BaU (as in Powell & Fuchs, 2010), but also for the contrast between the word-problem intervention condition without the pre-algebraic component versus BaU and for the contrast between the two variants of word-problem intervention.

These alterations provided a more stringent and informative test for assessing the paths by which word-problem intervention, with and without a pre-algebraic reasoning component, improved word-problem outcomes. We hypothesized sequential mediation for the contrasts involving word-problem intervention with pre-algebraic reasoning instruction, but not for the contrast between the word-problem intervention condition without the pre-algebraic component versus BaU. Thus, our main research question was, Do the paths by which word-problem intervention improves word-problem performance differ for word-problem intervention with versus without embedded pre-algebraic reasoning instruction? We also asked, What is the effect of each variant of the word-problem intervention on *equal sign*, *open equations*, and *word-problem* outcomes? and Does the word-problem intervention variant with embedded pre-algebraic reasoning provide added value over the same word-problem intervention without the pre-algebraic reasoning component on *equal sign*, *open equations*, and *word-problem* outcomes?

Method

Context and Setting

After receiving approval from our university's Institutional Review Board and school district for conducting research in public schools, we recruited elementary schools from a large urban school district in the Southwest of the U.S. This public school district served over 80,000 students. On average, the district reported 55.5% of students as Hispanic, 29.6% as Caucasian, 7.1% as African American, and 7.7% as belonging to another race or ethnic category. Overall, 27.1% of students identified as dual-language learners, 52.4% qualified as economically disadvantaged, and 12.1% received special education services. The district's graduation rate was 90.7%.

Participants

We recruited two cohorts of students for project participation across two years. During the 2016-2017 school year, for cohort 1, we recruited 37 third-grade teachers from 13 elementary schools. Several schools used departmentalization (i.e., the same teacher taught multiple mathematics classes), which accounted for the different numbers of teachers and classes. These 37 third-grade teachers taught 52 separate mathematics classes. From these 52 classrooms, we screened 916 third-grade students. During the 2017-2018 school year, for cohort 2, we recruited 44 teachers from 13 schools who taught 51 classrooms of students. We screened 818 third-grade students in the second cohort. In this study, we combined the data from cohorts 1 and 2 for a total of 1,734 third-grade students who participated in screening.

We screened all students using a measure of *Single-Digit Word Problems* (Jordan & Hanich, 2000). We used this measure to screen for mathematics difficulty (MD) in the area of word problems because word-problem solving was the primary focus of the intervention. For study eligibility, we identified students who answered 7 or fewer items correctly (out of 14) as experiencing MD. This cut-off score of 7 represented performance at or below the 25th percentile, a common cut-off score in research related to MD (Geary et al., 2012; Hecht & Vagi, 2010; Locuniak & Jordan, 2008). Based on the initial screening, we identified 472 students with MD. Of these, we did not pretest students for the following reasons: no parent consent or student assent (n = 28); Limited English proficiency of student (i.e., student experienced great difficulty with testing in English and teacher agreed student was not ready to participate in word-problem intervention; n = 49); student had a disability and received too many other services (n = 9); student moved before pretesting finished (n = 7); student had behavior issues identified by teacher (n = 12); teacher had too many students with MD in one classroom (n = 35), in this case, we randomly selected four students from each classroom to participate; student's parent opted the student out of study (n = 7); or could not schedule pretesting (n = 21). After completion of the pretest battery, we identified 304 third-grade students with MD across the two cohorts (see Table 1). The majority were female (56.3%), Hispanic (69.4%), and dual-language learners (60.5%). In addition, 12.2% of the students received special education services.

Random Assignment and Attrition

We randomly assigned the 304 students to three conditions: Pirate Math Equation Quest (PMEQ) intervention; Pirate Math without Equation Quest (PM-alone) intervention; and business-as-usual (BaU) comparison. Students were allocated to condition within teachers, and teachers were nested in schools. Overall, 20 students (6.7% of the 304 randomized students) did not complete the intervention because they (a) left the participating school prior to treatment's end (n = 16), (b) were discontinued from intervention due to disruptive behavior (n = 2), or (c) went into protective custody (n = 2). Attrition rates varied across treatment conditions. In BaU, 1% of students did not complete the posttest battery, while 6% of students in the PM-alone condition did not complete the intervention and or posttesting, and 15% of students in PMEQ failed to finish the intervention or complete posttesting. An overall attrition rate of 6.7% combined with differential attrition of 14% represents bias, which poses a threat to the study's internal validity. However, despite the differential attrition, the three groups were very similar on important baseline characteristics (Hedges' *g* ranged from 0.01 to 0.11; *p*-values ranged from .25 to .97; see Tables 1 and 2), suggesting equivalence prior to onset of treatment (What Works Clearinghouse, 2017). There was no attrition among teachers or schools.

General Education Instruction

All students with MD participated in regular mathematics instruction provided by their general education teacher. In the district, teachers primarily used the *GO Math!*, *Investigations in Number, Data, and Space for the Common Core*, or *Motivation Math* curricula to guide mathematics instruction. Students in the PMEQ and PM-alone conditions also received supplemental, individual intervention about word-problem solving. The interventionists did not provide intervention during the students' regular mathematics instruction to ensure students continued to fully participate in the district's mathematics curriculum.

Supplemental Intervention

For the students assigned to the PMEQ and PM-alone conditions, interventionists conducted sessions three times per week (i.e., 45 completed sessions) for 30 min a session. We aimed for students to complete 51 intervention sessions, but after accounting for student absences and schoolwide testing, we considered students who completed 45 to 51 sessions as completing the entire intervention. The interventionists worked with students individually in a quiet place outside of the classroom (e.g., school library, conference room, extra classroom). We assigned interventionists to work with students in both PMEQ and PM-alone groups to ensure an even quality of interventionists between the two intervention conditions.

PMEQ and PM-alone students participated in five activities for each session: (1) Math Fact Flashcards, (2) Equation Quest or Pirate Crunch, (3) Buccaneer Problems, (4) Shipshape Sorting, and (5) Jolly Roger Review. Only one activity (i.e., Equation Quest or Pirate Crunch) differed for students in the two intervention conditions. In the following sections, we describe each of the five activities in greater detail.

Math fact flashcards. To help increase math fact fluency, interventionists displayed a set of math fact flashcards (addends 0 to 9; minuends 0 to 18; and subtrahends 0 to 9) to students during two, 1-min timings. After setting the timer, students answered as many flashcards as they could in 1 min. Interventionists placed cards with a correct response on the desk and provided immediate, corrective feedback for incorrectly answered cards. After 1 min, interventionists and students counted the number of flashcards answered correctly. Prior to starting the second 1-min timing, interventionists challenged students to beat their previous score. At the end of the second 1-min timing, students graphed the highest score from the two trials.

Equation Quest. For PMEQ students only, Equation Quest served as the second activity of each intervention session. For approximately 2 to 5 min each session, interventionists provided instruction on solving equations and the meaning of the equal sign. Interventionists reintroduced the common symbol and taught students to understand the meaning of the equal sign as *the same as*. Students learned the equal sign acts as a balance between two sides of an equation and does not solely signal a calculation. To understand the equal sign as a relational symbol, students solved standard and nonstandard equations with concrete manipulatives (e.g., balance scale and blocks), hand-drawn pictures, or equations presented with numbers and symbols. Students learned a set of steps to balance equations with a variable (i.e., “X”), which involved isolating the variable and emphasizing that the calculation performed on one side of the equal sign also is performed on the other side of the equal sign (e.g., subtract 4 from both sides). Students practiced isolating the variable with both standard and nonstandard equations. For all PMEQ students, interventionists emphasized the meaning of the equal sign as *the same as* and embedded equation solving throughout each session.

Pirate Crunch. For PM-alone students only, Pirate Crunch was the second activity of each intervention session. For approximately 2 to 5 min each session, interventionists provided a mathematical review activity for students to complete. Pirate Crunch activities addressed concepts of telling time, money, geometry, perimeter, area, place value, and fractions through pencil and paper tasks. Pirate Crunch did not address concepts related to the equal sign or word problems.

Buccaneer Problems. The third activity for each session consisted of interventionist-led schema instruction through a series of three Buccaneer Problems. Note that PMEQ and PM-alone students received identical Buccaneer Problems. During sessions 1 through 4, interventionists reviewed addition and subtraction skills. Starting in session 5, interventionists provided explicit, scaffolded instruction on how to set up and solve word problems by schema. To emphasize new concepts, interventionists used manipulatives or pictures as necessary. Students learned to approach any word problem using the RUN attack strategy: *R*ead the problem, *U*nderline the label and cross out irrelevant information, and *N*ame the problem type (i.e., choose the correct schema to use). For each schema, students learned to use an equation to represent the problem and to mark “X” to represent the missing information. For the young pirates, “X” represented the treasure (i.e., a word-problem answer). Interventionists introduced the Total schema during session 5, the Difference schema in session 17, and the Change schema in session 34. From session 39 until the end of intervention, Buccaneer Problems included a comprehensive review of Total, Difference, and Change problems.

Shipshape Sorting. The fourth activity in each session, Shipshape Sorting, allowed students to practice identifying word-problem schemas learned during the Buccaneer Problems. Shipshape Sorting started during session 7 of the intervention. Before the sorting activity began, the interventionist placed a mat with four squares in front of the student. Each square was labeled with one word-problem type letter (i.e., T for Total, D for Difference, or C for Change) or a question mark. Interventionists reminded students to sort the word-problem cards and to not solve any of the word problems. Interventionists set the timer for 1 min and read the first word-problem card aloud before handing the card to the students. After 1 min, interventionists provided immediate, corrective feedback by reviewing at least three of the word-problem cards.

Jolly Roger Review. The final activity of each session, the Jolly Roger Review, included a brief, timed paper-and-pencil review of the session content. Students worked for 1 min to answer math facts, solve computation problems, or write appropriate equations for the three word-problem schemas. Then, students worked for 2 min to solve a word problem using the schema steps taught during the Buccaneer Problems. Students performed the timed review independently and then received feedback at the end of the 3 min.

Motivation. Throughout each session, students earned pirate coins for following specific guidelines such as listening to the interventionist, staying seated, working hard, and trying their best. Students typically earned 4 to 6 coins per session. At the end of each session, students counted the number of coins earned and colored the appropriate number of coins on a treasure map. When students completed a treasure map, they selected a small novelty prize from a treasure box.

Business-as-Usual Comparison

Students in the BaU condition did not receive any intervention from our research team. These students received regular classroom mathematics instruction. Classroom word-problem instruction for students in the BaU condition (as well as for students in the PMEQ and PM-alone conditions) incorporated general mnemonic devices (e.g., RICE: Read and restate, Illustrate, Calculate, Explain and edit), key word clues (e.g., altogether means add), and practice in applying problem-solution rules, as self-reported by participating teachers. Notably, none of the core mathematics classroom practices included schema instruction or explicit discussions about the equal sign as a relational symbol.

Examiners/Interventionists

We recruited 27 research assistants to act as examiners for pre- and posttesting and interventionists during tutoring. All research staff were pursuing or had obtained a Master's or doctoral degree in an education-related field. During the 2016-2017 school year (cohort 1), research staff (n = 15) were predominately female (n = 13), with 53% identifying as Caucasian (n = 8), 27% as Hispanic (n = 4), 13% as Asian American (n = 2), and 7% as African American (n = 1). During the 2017-2018 school year (cohort 2), all research staff were female (n = 15), with 73% (n = 11) identifying as Caucasian, 13% percent as Hispanic (n = 2), 7% as American Indian (n = 1), and 7% as African American (n = 1). Only 10% (n = 3) of research staff were the same from cohort 1 to cohort 2.

Throughout the year, research staff participated in trainings to ensure strong preparation for all aspects of the intervention. In late August and early September, examiners participated in three, 3-hr pretesting trainings. In early October, the team participated in two, 1.5-hr tutoring trainings about the content of the intervention and Total problems. Two subsequent 1.5-hr tutoring trainings followed in November to introduce Difference problems and in January to introduce Change problems. Lastly, examiners participated in one, 1.5-hr posttesting training meeting.

Fidelity of Implementation

We collected fidelity of implementation in several ways. First, for pretesting and posttesting, interventionists recorded all testing sessions. We randomly selected 472 of 2,366 (19.9%) audio recordings for analysis, evenly distributed across interventionists, and measured fidelity to testing procedures against detailed fidelity checklists. We measured pretesting fidelity at 98.5% (*SD* = 0.024) and posttesting fidelity at 98.8% (*SD* = 0.031).

Second, we measured fidelity of implementation of the interventions. The Project Manager conducted in-person fidelity observations once every three weeks for every interventionist. We also measured fidelity of intervention implementation through analysis of audio-recorded sessions. We audio-recorded every intervention session and randomly selected 1,632 of 8,160 (20.0%) audio-recorded sessions for analysis, evenly distributed across interventionists. Fidelity averaged 98% (*SD* = 0.037) for in-person supervisory observations and 98% (*SD* = 0.03) for audio-recorded intervention sessions.

Third, all interventionists tracked the number of sessions for their PMEQ and PM-alone students. We designed the intervention for students to finish at least 45 sessions with a maximum number of 51 sessions. The average PMEQ student completed 47.7 sessions of intervention (range 41 to 50; *SD* = 1.2), and the average PM-alone student completed 47.4 sessions of intervention (range 38 to 50; *SD* = 1.9). All sessions lasted 30 min, meaning PMEQ students received approximately 23.8 hours of intervention and PM-alone students received approximately 23.7 hours of intervention.

Internal Validity

As described, students in the BaU condition did not receive the word-problem intervention from our research team. To control for interventionists providing intervention to students in both active word-problem conditions, interventionists used separate color-coded packets of materials for the PMEQ and PM-alone students. During the initial trainings, we emphasized the importance of only using PMEQ strategies with students in the PMEQ condition and vice versa for PM-alone students. The lesson guides also included separate script sections for PMEQ and PM-alone students. Lastly, the in-person fidelity observations by the Project Manager ensured interventionists used the appropriate materials for each student. We did not identify any crossover mistakes during our in-person fidelity observations.

Measures

Screening and pretest measures. We used *Single-Digit Word Problems* as the primary measure for identifying students with MD (Jordan & Hanich, 2000). We administered this measure as the first in the whole-class pretesting session. *Single-Digit Word Problems* included 14 one-step word problems involving sums or minuends of 9 or less categorized into the Total, Difference, and Change schemas. In the U.S., standards outline that students should be able to set up and solve addition and subtraction (i.e., Total, Difference, and Change) word problems by the end of second grade (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Therefore, we interpreted low performance (i.e., <25th percentile with 7 or fewer items answered correctly) on *Single-Digit Word Problems* at the beginning of third grade as an indicator of MD. Examiners read each word problem aloud and could re-read each problem up to one time upon student request. We scored *Single-Digit Word Problems* as the number of correct responses (maximum = 14). We calculated Cronbach's α for the full screening sample of 1,734 third-grade students as .88.

During the whole-class pretesting session, we also administered *Texas Word Problems-Brief* (Powell & Berry, 2015). This measure included eight word problems requiring double-digit computation, with one Total, three Difference, and four Change problems respectively. For each problem, examiners read the problem aloud and provided approximately 1 min for students to solve the problem and write an answer. Examiners could re-read each problem up to one time upon student request. We scored this measure as the number of correct numerical and label responses for a maximum score of 16. Cronbach's α for the sample of students with MD was .83.

The next measure administered in the whole-class pretesting session was *Open Equations* (Powell, 2007). For *Open Equations*, students solved 10 equations in a standard (e.g., 3 + __ = 8) format. Students also solved equations in nonstandard formats, including two identity statements (e.g., __ = 4), 10 nonstandard equations with an operator symbol on the right side (e.g., 5 = 9 – __), and eight nonstandard equations with operator symbols on both sides (e.g., 9 – 6 = 7 – __). Excluding the identity statements, 14 of the equations included addition operator symbols and 14 included subtraction operator symbols. Students completed as many problems as possible within the 6-min timing. We scored this measure as the number of correct answers, with a maximum score of 30. Cronbach's α for students with MD was .86.

We administered *Equal Sign Tasks* as the fourth measure in the whole-class pretesting (Matthews & Rittle-Johnson, 2009). *Equal Sign Tasks* assessed students' understanding of the equal sign and equivalence in written format. First, examiners asked the students to write a definition of the equal sign. Next, students decided if the equal sign was used correctly in nonstandard closed equations. Then, students read statements of equivalence and decided whether each statement was always true, sometimes true, or never true. Finally, students viewed a closed equation with addends on both sides, broke the equation into two parts, and defined the meaning of the equal sign in the equation. The maximum score was 14. Cronbach's α for students with MD was .64.

Although we administered other measures across the three individual pretest sessions, we only discuss the measures used in the present report. In the first individual pretest session, we administered *Texas Word Problem-Part 1* (Powell & Berry, 2015). Students solved nine double-digit word problems: two Total problems, one Difference problem, four Change problems, and two multi-schema problems (i.e., Difference and Change; Total and Difference). Two problems featured the interpretation of graphs. Examiners read each problem aloud and provided students time to solve the problem and write an answer. Examiners could re-read each problem up to one time upon student request. We scored this measure as the number of correct numerical and label responses, with a maximum score of 18. Cronbach's α for students with MD was .84.

In the second individual pretesting session, examiners administered *Texas Word Problems-Part 2* (Powell & Berry, 2015). Students solved nine double-digit word problems: two Total problems, two Difference problems, three Change problems, one multi-schema problem (i.e., Total and Change), and one multiplicative problem (i.e., Equal Groups schema). Three problems featured the interpretation of graphs, and one problem included irrelevant information. The maximum score was 18. Cronbach's α for students with MD was .85.

Posttest measures. Examiners conducted five, 45-min posttesting sessions with groups of four or fewer students. Examiners administered *Open Equations* and *Texas Word Problems-Brief* in the first session. In session two, students completed *Texas Word Problems-Part 1* and *Equal Sign Tasks*. Examiners administered *Texas Word Problems-Part 2* in session three. All examiners followed identical procedures established during pretesting.

Scoring. Two examiners independently entered scores on 100% on the test protocols for each outcome measure on an item-by-item basis into an electronic database, resulting in two separate databases. We removed student names from all tests so examiners did not know the identify or condition of any student. We compared the discrepancies between the two databases across each outcome measure and rectified any inconsistencies to reflect the original response. Two examiners and the Project Manager resolved all discrepancies. Then, we converted students' responses to correct (1) and incorrect (0) scores using spreadsheet commands, which ensured 100% accuracy of scoring. Original scoring reliability was 99.9% for pretesting and 99.8% for posttesting.

For our analysis, we created a composite score of *Texas Word Problems-Brief*, *Texas Word Problems-Part 1*, and *Texas Word Problems-Part 2* by calculating the sum of the total score from each of the three measures at pretest or posttest. For this composite *word problems* score, we calculated Cronbach's α for students with MD at .94.

Procedure

During the first week of September, examiners administered whole-class pretesting in one, 55-min session. Identification of students with MD occurred shortly thereafter, with four weeks of individual pretesting during the last two weeks of September and the first two weeks of October. At pretest, examiners did not know which students had been assigned to treatment or BaU. During the third week of October, approximately 4 to 6 days after pretesting, intervention began and occurred three times per week for 16 weeks, concluding the third week in March. During this time, examiners filled the roles of interventionists. Approximately 4 to 6 days after the last intervention session, posttesting occurred in five, 45-min small group sessions with four students or fewer. We administered posttesting over three weeks, beginning the last week of March and ending the second week of April. We pre- and posttested all BaU students in the same time frame as the intervention students and in the same small groups as the intervention students.

Research Design and Nesting

We assigned students in the two active treatments (i.e., PMEQ or PM-alone) to an interventionist for purposes of delivering the intervention sessions. Placement with an interventionist was driven by students' existing and likely school schedules and by interventionists' availability; students were not randomized or "tracked" into interventionists nor were existing interventionists randomized to treatment conditions. Students randomized to BaU were not assigned to interventionists because BaU students did not receive supplemental intervention from the research team. This arrangement, where only a subset of a multilevel sample is nested (nested in interventionists here), is commonly described as a *partially nested randomized design* (Lohr et al., 2014). When units are randomized within blocks (teachers in this case), the design is a blocked partially nested randomized design. The nesting is *partial* because only a subset of students is subject to a given layer of nesting. All students in our study were nested in teachers (and teachers in schools), but, only students in the two treatment conditions *also* were nested in interventionists. Additionally, when comparing more than two conditions, the design can be described as a multi-arm blocked partially nested design. In the multi-arm scenario, some reasons favor random allocation at the partially nested level (randomized assignment to interventionists; see Lohr et al. 2014). However, randomizing to interventionists in school settings is simply not feasible in most cases (in our experience). That said, assignment to interventionists was systematic (non-random) only to the extent that placing students according to student and interventionist scheduled openings introduced patterned data (i.e., relatively unlikely).

Ignoring the asymmetry suggested by the partial nesting of students in interventionists is problematic because the different data structures for the treatment and the BaU groups imply different variance components. In our case, outcomes for treatment-assigned cases may vary by interventionists, unlike students assigned to BaU. We modeled the effect of interventionists in the treatment conditions only, following recommendations of Bauer et al. (2008) and Lohr et al. (2014). We modeled this effect as random, and allowed different variance estimates for treatment groups and BaU under the assumption that errors are independent.

These data were cross-classified in addition to being partially nested. In multilevel data, cross-classification occurs when cases from different levels of the model are not completely nested. For example, in our case, teacher and interventionist were crossed because students from the same teacher may have worked with different interventionists and because students from different teachers may have worked with the same interventionist (similar to Fuchs, Schumacher, et al., 2014). Note as well that cross-classification only occurs in the two treatment conditions because interventionists did not interact with students in the BaU. In this partially cross-classified structure (Luo et al., 2015), cases in one condition are nested under one random factor whereas cases in the other conditions are cross-classified across two random factors. The nested and cross-classified model components differ in their random effects. The random effect of interventionists exists only for students in one of the two treatment conditions, conditional on teacher effects. We modeled the data accordingly. Treatment-assigned cases were crossed on teacher and interventionist; BaU-assigned students were nested in teachers (see Luo et al., 2015).

Results

Preliminary Analyses

Table 2 displays means and standard deviations for each measure. All variables were distributed normally based on estimates of skewness and kurtosis, and we identified no outlying values. The groups were comparable at pretest, as described earlier. For our analysis, we refer to the score from *Equal Sign Tasks* as *equal sign* and the score on *Open Equations* as *open equations*. We refer to the composite score of the three word-problem measures as *word problems.*

Main Treatment Effects

Intervention conditions versus BaU. We treated the data as blocked, partially nested, and cross-classified, with students nested in teachers in all three conditions, students nested in interventionists in the two treated conditions (PMEQ and PM-alone), and interventionists and teachers crossed in the PMEQ and PM-alone groups. We estimated average treatment effects (ATE) in R using the *lme4* package (Luo et al., 2015) under intent-to-treat assumptions. We included scores at pretest for each outcome as covariates in the respective models. We calculated treatment effects as the across-arm mean difference divided by the *SD* within the control arm only (Lai & Kwok, 2016). Intraclass correlations (ICCs) at the teacher level were .05, .07, and .02 for *equal sign, open equations,* and *word problems*, respectively. Interventionist-level ICCs in the treatment groups were .15, .15, and .55 for *equal sign, open equations,* and *word problems*.

On *equal sign*, students in PMEQ outperformed students in BaU (β = 2.11, p = .00). The effect size was 0.73. The difference between PM-alone intervention and BaU was not significant (β = -0.75, p > .05, ES = - 0.26). On *open equations*, students in PM-alone outperformed students in BaU (β = 1.74, p < .001). The ES was 0.30. The contrast of PMEQ and BaU did not differ from 0 (β =.85, p > .05, ES = 0.14). On *word problems*, students in the PMEQ and PM-alone conditions significantly outperformed students in the BaU at posttest (β = 16.17, p < .001 and β = 14.86, p < .001, respectively). ESs for both contrasts were large: 2.66 for PMEQ versus BaU; 2.44 for the PM-alone versus BaU.

PMEQ versus PM-alone. We evaluated differences between the two treatment conditions as cross-classified models in R (Luo et al., 2015). On *equal sign*, students in PMEQ outperformed students in PM-alone (β = 2.91, p < .001). The ES was 1.01. On *open equations*, the contrast of PMEQ and PM-alone did not differ from 0 (β = -.67, p > .05, ES = -0.11). On *word problems*, students in the PMEQ and PM-alone conditions were not significantly different from one another (β = 1.31, p > .05, ES = 0.22).

Mediation Analysis

We modeled the mediating effect of *equal sign* and *open equations* using multilevel path analytic framework (Bauer, Preacher & Gil, 2006). We estimated three indirect effects, all involving the path through *equal sign* and *open equations* (see Figure 1) but differing in the groups being contrasted. These models represent sequential mediation where treatment causes improved performance on *equal sign*, which covaries with improved scores on *open equations*, which in turn covaries with improved scores on *word problems.* We describe the latter two paths as covariances because levels of treatment were not re-randomized at each step in the sequence of relations beyond the initial effect on *open equations*. However, given the model's strong theoretical support and following the logic of Pearl (2009, 2014) and others (Imai et al., 2010), it is reasonable to conceptualize the indirect effect as a "unitary" parameter to which students are randomized, allowing for causal inferences across the arc of effects in Figure 1. Pearl (2014) demonstrated the assumptions for mediation often cited in the methodological literature are overly restrictive (e.g., randomized mediators within strata of the independent variable) and can be relaxed substantially without compromising identification. "Natural effects" can be identified even when treatment assignment remains formally confounded with the mediator or with the outcome, particularly in parametric models including interactions, allowing comparisons of the relative importance of several mediated pathways without bias.

All variables in the model were measured at level 1 (student level). Treatment's effect on *equal sign* (path a_1 and a_2) and *word problems* (paths c'_1 and c'_2) and the effect of *equal sign* on *open equations* (path *b*) and *word problem* on *equal sign* (path *c*) were allowed to vary across interventionist units in the two treatment groups. We estimated each indirect effect and its related direct effect simultaneously and established confidence intervals for indirect effects using Monte Carlo simulations with 100,000 iterations (Bauer et al., 2006; Preacher & Selig, 2012). This required "stacking" the data (based on recommendations of Bauer et al., 2006). In our case, data for the model with *equal sign* (*M1*) as the dependent variable and data for the model with *open equations* (*M2*) as a dependent variable were stacked with data for the model with *word problems* (*Y*) as the dependent variable (Bauer et al., 2006). This also required creation of an outcome variable (*Z*) containing stacked values for both *M1*, *M2* and *Y*. Dummy selector variables were included in the dataset corresponding to whether *Z* referenced values for *M1*, *M2,* or *Y* on each data line. The mediation effect was estimated as the product of the "a_1" or "a_2" arm of the model (or the relation between treatment condition and *equal sign*), the "*b*" arm of the model (or the relation between *equal sign* and *open equations*), and the "*c*" arm of the model (the relation between *open equations* and *word problems*), adding "σ_{ajbj}", "σ_{ajcj}", and "σ_{bjcj}", which are the covariances between the random effects for the "*a*" and "*b*", "*a*" and "*c*" , and "*b*" and "*c*" arms of the model (Bauer et al., 2006; Bolger & Laurenceau, 2013).

The indirect effect of PMEQ versus BaU on *word problems* was significant ($\beta = .50$, $p < .001$; Monte Carlo $CI_{95} = 0.10, 0.71$), indicating participation in the PMEQ intervention as opposed to BaU contributed to improved *word problems* performance via improving student performance on *equal sign* and *open equations*. By contrast, the indirect effect of PM-alone on *word problems* was -0.18 (CI_{95} = -0.38, 0.11), indicating a non-significant effect of PM-alone on *word problems* via *equal sign* and *open equations*. The indirect effect of PMEQ versus PM-alone on *word problems* was significant (β = .49, p = .00; Monte Carlo CI_{95} = 0.03, 0.98), meaning that participation in the PMEQ intervention as opposed to PM-alone improved *word problems* scores by improving students' performance on *equal sign* and *open equations*.

We also examined whether demographic variables acted as moderators of intervention effects. These regression models, based on the path analytic framework (Preacher & Hayes, 2008), were extensions of the main effect analyses. We expanded the statistical model to include the moderators as a predictor of differential response and its interaction with the condition effect. We evaluated three demographic variables as potential moderators including gender, special education status, and dual-language learner status. We identified no differences in the effectiveness of intervention (p-values ranged from .19 to .92) on *equal sign*, *open equations*, or *word problems*.

Discussion

In this study, we explored paths by which word-problem intervention, with versus without embedded pre-algebraic reasoning instruction, improves word-problem performance. We extended Powell and Fuchs's (2010) quasi-experiment in a variety of ways, including an expanded and redesigned pre-algebraic reasoning component as well as a focus on all three additive word-problem schemas used in the early elementary grades. The present study also included two important methodological extensions to Powell and Fuchs (2010), with a randomized controlled trial and elaborated mediation models, in which equal-sign understanding and equation solving were considered as sequential mediators and in which mediation was considered for all three study condition contrasts.

We hypothesized there would be sequential mediation for the two contrasts involving PMEQ (against BaU and against PM-alone), but not for the contrast between the PM-alone condition without the pre-algebraic component versus BaU. As defined by Pillay et al. (1998), pre-algebraic reasoning involves (a) interpretation of the equal sign as relational and (b) solving equations with one unknown, which motivated our inclusion of these two factors as potential mediators. Although Pillay et al. (1998) did not address the sequential development of these forms of pre-algebraic reasoning, other studies provide the basis for positing that interpretation of the equal sign as relational leads to improved equation solving (Knuth et al., 2006; Matthews et al., 2012; Rittle-Johnson & Alibali, 1999).

Thus, with sequential mediation, we hypothesized that participation in PMEQ, but not PM-alone, would produce improved *equal sign* understanding, which is associated with stronger *open equations* performance, which in turn is associated with stronger *word problem* outcomes. Results were consistent with our hypothesis. For contrasts involving the PMEQ condition, the indirect effect on word-problem outcomes, via equal-sign understanding and open equation solving, was significant.

In this way, the effects of PMEQ on word-problem outcomes accrued via a combination of direct and indirect effects: It directly improved word-problem performance (see the large direct effect, represented by the *c*' path in the lower triangular representation in Figure 1, even after the indirect effect was included in the model), and PMEQ indirectly strengthened word-problem solving via improved equal-sign understanding and equation solving. By contrast, the indirect effect on word-problem outcomes for the contrast of PM-alone versus BaU was not significant. The word-problem advantage for PM-alone over BaU instead accrued entirely directly (see its large direct effect on word-problem outcomes, represented by the *c*' path in in Figure 1's higher triangular representation).

In line with these mediation effects, our main effects models revealed that PMEQ students' posttest *equal sign* performance was superior to that of both BaU and PM-alone students, with large ESs of 0.73 and 1.01, respectively. These results reflect PMEQ's emphasis on the equal sign as a relational symbol, which included learning to interpret the equal sign as meaning *the same as*, using pictorial drawings or balance scales with hands-on manipulatives to experience the equal sign as relational, and participating in ongoing discussions about the relational definition of the equal sign. At the same time, posttest equal-sign performance for PM-alone students was comparable to that of BaU students (ES = -0.26). Consequently, although the act of generating equations to represent word problems (e.g., 5 + X = 13) and solving those open equations did not promote relational understanding of the equal sign, writing equations did improve open-equation solving, in line with Powell and Fuchs (2010).

Surprisingly, however, PM-alone but not PMEQ students demonstrated superior posttest nonstandard equation-solving performance relative to the BaU condition. The relatively small but significant ES comparing PM-alone to BaU students was 0.30, while the ES contrasting PMEQ and BaU students was 0.14. This finding was unexpected because PMEQ students received more elaborated equation-solving instruction as part of the pre-algebraic reasoning component. This instruction included teaching students to remove a constant from one side of an equation to isolate a variable ("X"). Notably, a major extension to equation instruction in the present study involved teaching students to isolate the unknown variable on one side of equation. This task likely challenged our students with MD, and future studies should examine whether the effects of such instruction at third grade are moderated by students' pretest mathematics skills or domain-general abilities. Such a moderator effect may have muted the overall effects of the algebraic reasoning component's effect on *open equations* in the PMEQ condition.

Importantly, the main effects model revealed superior *word problem* outcomes for PMEQ and for PM-alone students over BaU students, with large ESs (2.66 and 2.44). Moreover, the difference on word-problem outcomes did not differ between conditions. We attribute superior word-problem performance to the extensive instruction students in both active word-problem interventions received on interpreting, setting up, and solving word problems. Our word-problem intervention relied on schema instruction, in which students learned to identify a word problem as belonging to a specific type or schema and solve the word problem according to the learned schema strategy. These results corroborate prior research demonstrating the efficacy of schema instruction for students with MD (Fuchs et al., 2008; Fuchs, Powell, et al., 2014; Jitendra et al., 2015; Xin et al., 2011).

Limitations

Before offering our major conclusion, we note three study limitations. First, we only asked students to respond to pre- and posttest assessments in writing. That is, students answered all questions on *Equal Sign Tasks*, *Open Equations*, and the *Texas Word Problems* assessments using pencil and paper. Thus, we assumed written responses represented students' knowledge of the various forms of mathematics content. Future research should include assessments that capture students' oral responses to determine whether written responses accurately assess students' knowledge on these measures. Furthermore, each item on *Texas Word Problems* was read to the student by an examiner. Future research should investigate this accommodation and compare student performance when word problems are read silently by the student or read aloud by the examiner.

Second, we did not collect data to determine the extent to which interventionists working with PMEQ students emphasized Equation Quest components throughout other components of each lesson (e.g., Buccaneer Problems, Shipshape Sorting, or Jolly Roger Review). To understand the quantity and quality of review necessary for students to establish strong pre-algebraic reasoning skills, future studies should quantify how much pre-algebraic reasoning emphasis is deemed adequate for students with MD. Third, readers should note that differential attrition occurred among the study conditions, a problem that was mitigated in part by the fact that the study groups, after attrition, were comparable on all demographic and pretest performance variables.

Conclusion

We conclude that both variants of a word-problem intervention convey significant and large practical advantages over the BaU condition on word-problem outcomes for students with MD. Yet, while both conditions improved word-problem performance to a comparable and impressive extent, the paths by which word-problem effects accrued differed for the two variants of the intervention. Schema word-problem intervention without a pre-algebraic reasoning instruction component did not improve equal-sign understanding, and its effect on equation solving, although significant, was relatively small. This outcome precluded a significant indirect effect via pre-algebraic reasoning. By contrast, schema word-problem intervention with a pre-algebraic reasoning instruction component benefitted performance with large effects not only on word-problem performance but also on relational equal-sign understanding, which produced a significant indirect effect via pre-algebraic reasoning.

On this basis and given the link between equal-sign understanding and success with algebra, our research suggests the word-problem intervention condition with embedded pre-algebraic reasoning instruction offers a practical advantage. Thus, word-problem intervention condition with embedded pre-algebraic reasoning instruction represents an effective and efficient alternative to the contrasting word-problem intervention condition without a pre-algebraic reasoning instruction component.

References

Aunio, P., Heiskari, P., Van Luit, J. E. H., & Vuorio, J.-M. (2015). The development of early numeracy skills in kindergarten in low-, average- and high-performance groups. *Journal of Early Childhood Research, 13*(1), 3–16. https://doi.org/10.1177/1476718X14538722

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods, 11*(2), 142–163. https://doi.org/10.1037/1082-989X.11.2.142

Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research, 43*(2), 210–236. https://doi.org/10.1080/00273170802034810

Blanton, M., Stephens, A., Knuth, E., Gardiner, A. M., Isler, I., & Kim, J.-S. (2015). The development of children's algebraic thinking: The impact of a comprehensive early algebra intervention in third grade. *Journal for Research in Mathematics Education, 46*(1), 39–87. https://doi.org/10.5951/jresematheduc.46.1.0039

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* New York, NY: Guilford Press.

Branum-Martin, L., Fletcher, J. M., & Stuebing, K. K. (2012). Classification and identification of reading and math disabilities: The special case of comorbidity. *Journal of Learning Disabilities, 46*(6), 490–499. https://doi.org/10.1177/0022219412468767

Bryant, B. R., Bryant, D. P, Porterfield, J., Dennis, M. S., Falcomata, T., Valentine, C., …Bell, K. (2016). The effects of a tier 3 intervention on the mathematics performance of second grade students with severe mathematics difficulties. *Journal of Learning Disabilities,* *49*(2), 176–188. https://doi.org/10.1177/0022219414538516

Byun, S., Irvin, M. J., & Bell, B. A. (2015). Advanced math course taking: Effects on math achievement and college enrollment. *The Journal of Experimental Education, 83*(4), 439–468. https://doi.org/10.1080/00220973.2014.919570

Cook, S. C., Collins, L. W., Morin, L. L., & Riccomini, P. J. (2019). Schema-based instruction for mathematical word problem solving: An evidence-based review for students with learning disabilities. *Learning Disability Quarterly*. https://doi.org/10.1177/0731948718823080

Driver, M. K., & Powell, S. R. (2015). Symbolic and nonsymbolic equivalence tasks: The influence of symbols on students with mathematics difficulty. *Learning Disabilities Research and Practice, 30*(3), 127–134. https://doi.org/10.1111/ldrp.12059

Eddy, C. M., Fuentes, S. Q., Ward, E. K., Parker, Y. A., Cooper, S., Jasper, W. A., Mallam, W. A., Sorto, M. A., & Wilkerson, T. L. (2015). Unifying the algebra for all movement. *Journal of Advanced Academics, 26*(1), 59–92. https://doi.org/10.1177/1932202X14562393

Flores, M. M., Hinton, V. M., & Burton, M. E. (2016). Teaching problem solving to students receiving tiered interventions using the concrete-representational-abstract sequence and schema-based instruction. *Prevention School Failure: Alternative Education for Children and Youth, 60*(4), 345–355. https://doi.org/10.1080/1045988X.2016.1164117

Fuchs, L. S., Powell, S. R., Cirino, P. T., Schumacher, R. F., Marrin, S., Hamlett, C. L., Fuchs, D., Compton, D. L., & Changas, P. C. (2014). Does calculation or word-problem instruction provide a stronger route to pre-algebraic knowledge? *Journal of Educational Psychology, 106*(4), 990–1006. https://doi.org/10.1037/a0036793

Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, J., Malone, A., Hamlett, C. L., Jordan, N. C., Gersten, R., Siegler, R., & Changas, P. (2014). Does working memory moderate the effects of fraction intervention? An aptitude-treatment interaction. *Journal of Educational Psychology, 106*(2), 499–514. https://doi.org/10.1037/a0034341

Fuchs, L. S., Seethaler, P. M., Powell, S. R., Fuchs, D., Hamlett, C. L., & Fletcher, J. M. (2008). Effects of preventative tutoring on the mathematical problem solving of third-grade students with math and reading difficulties. *Exceptional Children, 74*(2), 155–173. https://doi.org/10.1177/001440290807400202

García, A. I., Jiménez, J. E., & Hess, S. (2006). Solving arithmetic word problems: An analysis of classification as a function of difficulty in children with and without arithmetic LD. *Journal of Learning Disabilities, 39*(3), 270–281. https://doi.org/10.1177/00222194060390030601

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

Griffin, C. C., Gagnon, J. C., Jossi, M. H., Ulrich, T. G., & Myers, J. A. (2018). Priming mathematics word problem structures in a rural elementary classroom. *Rural Special Education Quarterly, 37*(3), 150–163. https://doi.org/10.1177/8756870518772164

Griffin, C. C., & Jitendra, A. K. (2009). Word problem-solving instruction in inclusive third-grade mathematics classrooms. *The Journal of Educational Research, 102*(3), 187–201. https://doi.org/10.3200/joer.102.3.187-202

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling [White paper]. Retrieved from http://www.afhayes.com/ public/process2012.pdf

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

Hoffer, T. B., Venkataraman, L., Hedberg, E. C., & Shagle, S. (2007). *Final report on the national survey of algebra teachers for the National Math Panel.* Chicago, IL: National Opinion Research Center.

Imai, K., Keele, L., & Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science, 25*(1), 51–71. https://doi.org/10.1214/10-STS321

Jacobs, V. R., Franke, M. L., Carpenter, T. P., Levi, L., & Battey, D. (2007). Professional development focused on children's algebraic reasoning in elementary school. *Journal for Research in Mathematics Education, 38*(3), 258–288. https://doi.org/10.2307/30034868

Jitendra, A. K., Harwell, M. R., Dupuis, D. N., & Karl, S. R. (2017). A randomized trial of the effects of schema-based instruction on proportional problem-solving for students with mathematics problem-solving difficulties. *Journal of Learning Disabilities, 50*(3), 322–336. https://doi.org/10.1177/0022219416629646

Jitendra, A. K., Petersen-Brown, S., Lein, A. E., Zaslofsky, A. F., Kunkel, A. K., Jung, P.-G., & Egan, A. M. (2015). Teaching mathematical word problem solving: The quality of evidence for strategy instruction priming the problem structure. *Journal of Learning Disabilities, 48*(1), 51–72. https://doi.org/10.1177/0022219413487408

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

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*(1), 109–129. https://doi.org/10.1037//0033-295x.92.1.109

Knuth, E. J., Stephens, A. C., McNeil, N. M., & Alibali, M. W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education, 37*(4), 297–312.

Lai, M. H. C., & Kwok, O.-M. (2016). Estimating standardized effect sizes for two- and three-level partially nested data. *Multivariate Behavioral Research, 51*(6), 740–756. https://doi.org/10.1080/00273171.2016.1231606

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities*, *41*(5), 451–459. https://doi.org/10.1177/0022219408321126

Lohr, S., Schochet, P.Z., and Sanders, E. (2014). *Partially nested randomized controlled trials in education research: A guide to design and analysis.* (NCER 2014-2000) Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Luo, W., Cappaert, K. J., & Ning, L. (2015). Modelling partially cross-classified multilevel data. *British Journal of Mathematical and Statistical Psychology, 68*(2), 342–362. https://doi.org/10.1111/bmsp.12050

Martin, R. B., Cirino, P. T., Barnes, M. A., Ewing-Cobbs, L., Fuchs, L. S., Stuebing, K. K., & Fletcher, J. M. (2013). Prediction and stability of mathematics skill and difficulty.

Journal of Learning Disabilities, 46(5), 428–443. https://doi.org/10.1177/0022219411436214

Matthews, P., & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations, concepts, and procedures as pedagogical tools. *Journal of Experimental Child Psychology, 104*(1), 1–21. https://doi.org/10.1016/j.jecp.2008.08.004

Matthews, P., Rittle-Johnson, B., McEldoon, K., & Taylor, R. (2012). Measure for measure: What combining diverse measures reveals about children's understanding of the equal sign as an indicator of mathematical equality. *Journal for Research in Mathematics Education, 43*(3), 316–350.

McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development, 76*(4), 883–899. https://doi.org/10.1111/j.1467-8624.2005.00884.x

McNeil, N. M., Fyfe, E. R., Petersen, L. A., Dunwiddie, A. E., Brletic-Shipley, H. (2011). Benefits of practicing 4 = 2 + 2: Nontraditional problem formats facilitate children's understanding of mathematical equivalence. *Child Development, 82*(5), 1620–1633. https://doi.org/10.1111/j.1467-8624.2011.01622.x

National Governors Association Center for Best Practices & Council of Chief State School Officers (2010). *Common core state standards mathematics.* Washington, DC: Author.

Peake, C., Jiménez, J. E., Rodríguez, C., Bisschop, E., & Villarroel, R. (2015). Syntactic awareness and arithmetic word problem solving in children with and without learning disabilities. *Journal of Learning Disabilities, 48*(6), 593–601. https://doi.org/10.1177/0022219413520183

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). Cambridge, England: Cambridge University Press.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods, 19*(4), 459–481. https://doi.org/10.1037/a0036434

Peltier, C., & Vannest, K. J. (2017). A meta-analysis of schema instruction on the problem-solving performance of elementary school students. *Review of Educational Research, 87*(5), 899–920. https://doi.org/10.3102/0034654317720163

Pillay, H., Wilss, L., & Boulton-Lewis, G. (1998). Sequential development of algebra knowledge: A cognitive analysis. *Mathematics Education Research Journal, 10*(2), 87–102. https://doi.org/10.1007/bf03217344

Powell, S. R. (2007). *Open Equations.* Available from S. R. Powell, 1912 Speedway, D5300, Austin, TX 78712.

Powell, S. R. (2012). Equations and the equal sign in elementary mathematics textbooks. *The Elementary School Journal, 112*(4), 627–648. https://doi.org/10.1086/665009

Powell, S. R., & Berry, K. A. (2015). *Texas Word Problems.* Available from S. R. Powell, 1912 Speedway, D5300, Austin, TX 78712.

Powell, S. R., Driver, M. K., & Julian, T. E. (2015). The effect of tutoring with nonstandard equations for students with mathematics difficulty. *Journal of Learning Disabilities, 48*(5), 523–534. http://dx.doi.org/10.1177/0022219413512613

Powell, S. R., & Fuchs, L. S. (2010). Contribution of equal-sign instruction beyond word-problem tutoring for third-grade students with mathematics difficulty. *Journal of Educational Psychology, 102*(2), 381–394. https://doi.org/10.1037/a0018447

Powell, S. R., & Fuchs, L. S. (2014). Does early algebraic reasoning differ as a function of students' difficulty with calculations versus word problems? *Learning Disabilities Research and Practice, 29*(3), 106–116. https://doi.org/10.1111/ldrp.12037

Powell, S. R., Fuchs, L. S., Cirino, P. T., Fuchs, D., Compton, D. L., & Changas, P. C. (2015). Effects of a multitier support system on calculation, word problem, and pre-algebraic learning among at-risk learners. *Exceptional Children, 81*(4), 443–470. https://doi.org/10.1177/0014402914563702

Powell, S. R., Kearns, D. M., & Driver, M. K. (2016). Exploring the connection between arithmetic and pre-algebraic reasoning at first and second grade. *Journal of Educational Psychology, 108*(7), 943–959. https://doi.org/10.1037/edu0000112

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*(3), 879–891. https://doi.org/10.3758/BRM.40.3.879

Preacher, K. J., & Selig, J. P. (2012). Advantages of Monte Carlo confidence intervals for indirect effects. *Communication Methods and Measures, 6*(2), 77–98. https://doi.org/10.1080/19312458.2012.679848

Rittle-Johnson, S., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*(1), 175–189. https://doi.org/10.1037/0022-0663.91.1.175

Stephens, A. C., Knuth, E. J., Blanton, M. L., Isler, I., Gardiner, A. M., & Marum, T. (2013). Equation structure and the meaning of the equal sign: The impact of task selection in eliciting elementary students' understandings. *The Journal of Mathematical Behavior, 32*(2), 173–182. https://doi.org/10.1016/j.jmathb.2013.02.001

Swanson, H. L., Orosco, M. J., & Lussier, C. M. (2014). The effects of mathematics strategy instruction for children with serious problem-solving difficulties. *Exceptional Children, 80*(2), 149–168. https://doi.org/10.1177/001440291408000202

Tolar, T. D., Fuchs, L., Fletcher, J. M., Fuchs, D., & Hamlett, C. L. (2016). Cognitive profiles of mathematical problem solving learning disability for different definitions of disability. *Journal of Learning Disabilities, 49*(3), 240–256. https://doi.org/10.1177/0022219414538520

van Garderen, D., Scheuermann, A., & Jackson, C. (2012). Examining how students with diverse abilities use diagrams to solve mathematics word problems. *Learning Disability Quarterly, 36*(3), 145–160. https://doi.org/10.1177/0731948712438558

Vincent, J., Bardini, C., Pierce, R., & Pearn, C. (2015). Misuse of the equals sign: An entrenched practice from early primary years to tertiary mathematics. *Australian Senior Mathematics Journal, 29*(2), 31–39.

What Works Clearinghouse (2017). *What Works Clearinghouse: Procedures and Standards Handbook, Version 4.0.* Washington, DC: Author. Retrieved from: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Willis, G. B., & Fuson, K. C. (1988). Teaching children to use schematic drawings to solve addition and subtraction word problems. *Journal of Educational Psychology, 80*(2), 192–201. http://dx.doi.org/10.1037/0022-0663.80.2.192

Xin, Y. P., Whipple, A., Zhang, D., Si, L., Park, J. Y., & Tom, K. (2011). A comparison of two mathematics problem-solving strategies: Facilitate algebra-readiness. *The Journal of Educational Research, 104*(6), 381–395. https://doi.org/10.1080/00220671.2010.487080

Xin, Y. P., & Zhang, D. (2009). Exploring a conceptual model-based approach to teaching situated word problems. *The Journal of Educational Research, 102*(6), 427–441. https://doi.org/10.3200/joer.102.6.427-442

Zheng, X., Flynn, L. J., & Swanson, H. L. (2012). Experimental intervention studies on word problem solving and math disabilities: A selective analysis of the literature. *Learning Disability Quarterly, 36*(2), 97–111. https://doi.org/10.1177/0731948712444277

Table 1

Participant Demographics

	Overall (N = 304)	PMEQ (n = 105)	PM-alone (n = 84)	BaU (n = 115)
Gender (% female)	56.3	59.0	50.0	58.3
Race/ethnicity				
African-American	11.5	12.4	11.9	10.4
Asian	2.6	1.9	3.6	2.6
Hispanic/Latinx	69.4	62.9	75.0	71.3
Caucasian	5.3	6.7	3.6	5.2
Multi-racial	6.6	8.6	4.8	6.1
Other	2.3	2.9	0.0	3.5
Students in special education	12.2	11.4	16.7	9.6
Dual-language learners	60.5	61.9	60.7	59.1

Note. BaU = Business as usual; PM-alone = Pirate Math without Equation Quest; PMEQ = Pirate Math Equation Quest.

Table 2

Pretest and Posttest Means with Standard Deviations for Outcome Measures

		Pretest			Posttest		
		n	*M*	*SD*	*n*	*M*	*SD*
Equal sign	PMEQ	105	5.64	2.70	91	10.12	2.96
	PM-alone	84	5.74	2.77	79	7.16	2.78
	BaU	115	5.70	2.79	114	7.87	2.88
Open equations	PMEQ	105	5.40	3.63	91	11.49	6.04
	PM-alone	84	5.17	4.46	79	12.13	4.74
	BaU	115	5.84	4.26	114	10.61	5.86
Word problems	PMEQ	105	6.93	5.11	91	27.52	10.59
	PM-alone	84	6.60	4.94	79	25.30	10.23
	BaU	115	6.61	5.00	114	10.67	6.08

Note. BaU = Business as usual; PMEQ = Pirate Math Equation Quest; PM-alone = Pirate Math without Equation Quest.

Table 3

Results from Cross-Classified Partially Nested Models Testing Effects of Intervention

	Equal sign			*Open equations*			*Word problems*		
Fixed effects	Estimate (*SE*)		Effect size	Estimate (*SE*)		Effect size	Estimate (*SE*)		Effect size
Intercept	7.94 (0.26)	***		10.52 (0.51)	***		10.69 (0.68)	***	
Pretest	0.37 (0.06)	***		0.55 (0.07)	***		0.62 (0.09)	***	
PM-alone vs BAU	-0.75 (0.41)		-0.26	1.74 (0.72)	*	0.30	14.86 (1.53)	***	2.44
PMEQ vs BAU	2.11 (0.45)	***	0.73	0.85 (0.86)		0.14	16.17 (1.61)	***	2.66
Random effects	Variance		*SD*	Variance		*SD*	Variance		*SD*
Student-level									
Intercept	5.48		2.34	20.93		4.58	38.55		6.23
Tutor-level									
Intercept	0.54		0.88	6.20		2.49	2.47		1.96
PM-alone vs BAU	0.79		0.46	6.06		2.46	18.74		5.37
PMEQ vs BAU	0.18		0.87	0.00		2.87	23.96		5.01
Teacher-level									
Intercept	0.48		0.69	0.95		0.97	3.50		1.87

Note. BAU = Business as usual; PM-alone = Pirate Math without Equation Quest; PMEQ = Pirate Math Equation Quest.

***p > .001; *p > 0.05

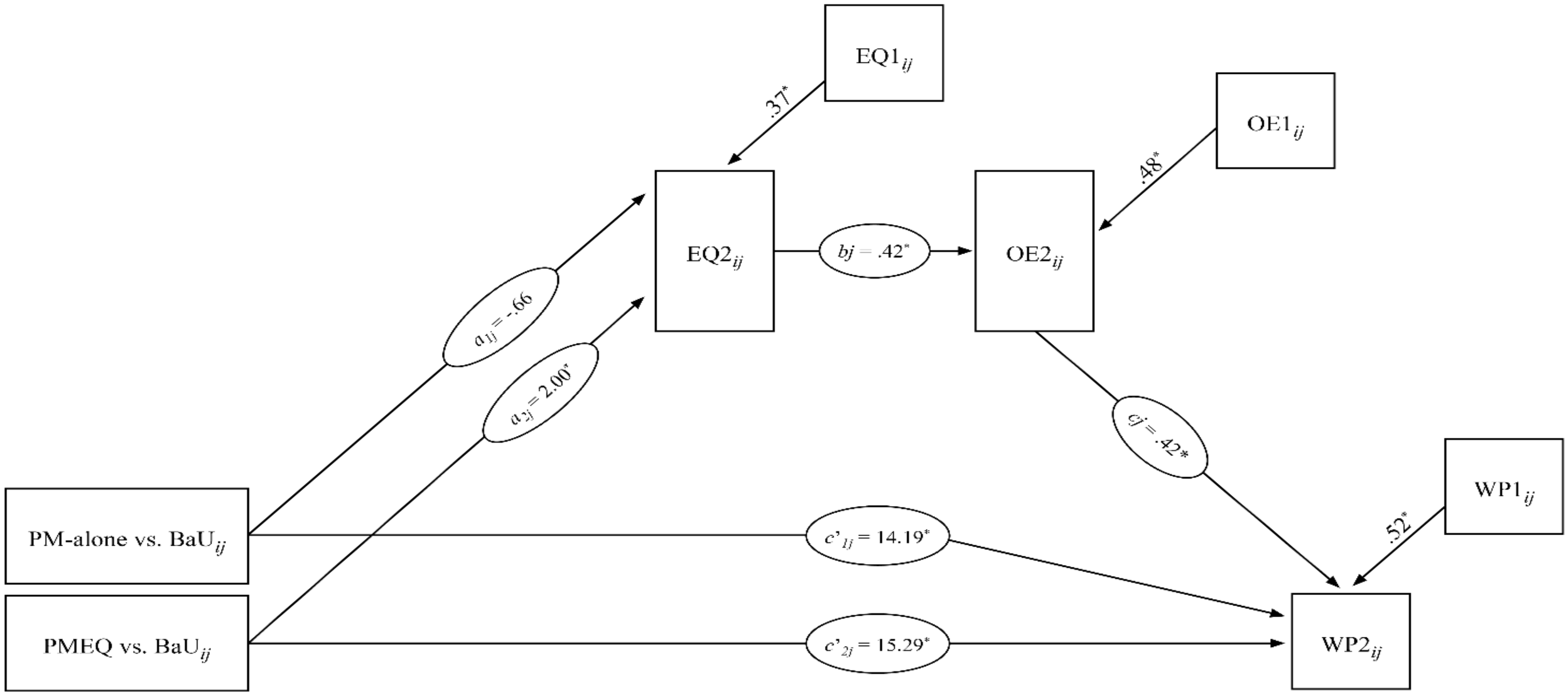

Figure 1. Multilevel mediation model for the 1-1-1-1 design.

Note. BaU = Business as usual; EQ1 = *Equal sign* pretest; EQ2 = *Equal sign* posttest; OE1 = *Open equations* pretest; OE2 = *Open equations* posttest; PM-alone = Pirate Math without Equation Quest; PMEQ = Pirate Math Equation Quest; WP1 = *Word problems* pretest; WP2 = *Word Problems* posttest. Subscript *j* represents interventionists. As indicated in the manuscript, we allowed each path in the system to differ between interventionists. a_j, b_j, and c_j represent the random effects.