Applying an Individual Word-Problem Intervention to a Small-Group Setting: A Pilot Study's Evidence

Sarah R. Powell, Katherine A. Berry, Anasazi N. Acunto, Anna-Maria Fall, and Greg Roberts

The University of Texas at Austin

Powell, S. R., Berry, K. A., Acunto, A. N., Fall, A.-M., & Roberts, G. (2022). Applying an individual word-problem intervention to a small-group setting: A pilot study's evidence of improved word-problem performance for students experiencing mathematics difficulty. *Journal of Learning Disabilities*, *55*(5), 359–374. https://doi.org/10.1177/00222194211047635

Author Note

Sarah R. Powell https://orcid.org/0000-0002-6424-6160

Katherine A. Berry https://orcid.org/0000-0001-8340-3259

Anasazi N. Acunto

Anna-Maria Fall https://orcid.org/0000-0002-6257-6684

Greg Roberts https://orcid.org/0000-0001-6333-7442

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@utexas.edu

Abstract

The purpose of this pilot study was to determine whether positive results from a word-problem intervention implemented one-to-one contributed to similar outcomes when implemented in small groups of three to four students. Third-grade students experiencing mathematics difficulty (n = 76) were randomly assigned to word-problem intervention (n = 56) or business-as-usual comparison (n = 20). Intervention occurred for 13 weeks, 3 times per week, 30 min per session. Multilevel models revealed the intervention condition significantly outperformed the BaU on a proximal word-problem outcome, corroborating results from our prior individual intervention. When comparing student performance in the individual versus small-group intervention, findings suggest students received added benefit from the individual intervention. The word-problem intervention successfully translated to a small-group setting, which holds important implications for educators working with students in supplemental, targeted, or Tier-2 mathematics intervention settings.

Key words: learning difficulty; mathematics; small group; word problems

Applying an Individual Word-Problem Intervention to a Small-Group Setting: A Pilot Study's Evidence

In the elementary grades, students must interpret and solve word problems to develop mathematics competency. 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., "Lily ran for 23 minutes during the track meet. Luis ran 16 minutes longer than Lily. How many minutes did Luis run during the track meet?"). In the United States, 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. In fact, the majority of mathematics items on high-stakes assessments are embedded within a word-problem scenario (Powell, Namkung, et al., in press); thus, word-problem proficiency proves essential for demonstrating successful mathematics performance.

Although word-problem solving is demanding for many students (Fuchs et al., 2014), students experiencing mathematics difficulty (MD) often demonstrate markedly poor performance relative to students without MD (Peake et al., 2015; van Garderen et al., 2012). Fortunately, word-problem outcomes for students experiencing MD can improve with word-problem intervention (Flores et al., 2016; Fuchs et al., 2014; Griffin et al., 2018; Swanson et al., 2014; Xin et al., 2011; Zheng et al., 2012). In this pilot study, we investigated whether a word-problem intervention designed to be implemented with individual students experiencing MD could be implemented with small groups of students experiencing MD and lead to positive word-problem outcomes. To provide the background for this study, in the introduction, we discuss students experiencing MD and their difficulties with mathematics. Then, we highlight word-problem intervention focused on schemas and research using this strategy with students experiencing MD. Next, we review the implementation of mathematics intervention with a focus on individual versus small-group implementation. Finally, we present the purpose and research questions of the present study.

Students Experiencing MD

Students identified with a learning disability in mathematics, sometimes referred to as dyscalculia (e.g., Butterworth, 2010), account for approximately 3 to 6% of all school-age students (Devine et al., 2018; Morsanyi et al., 2018; Shalev et al., 2000). Beyond disability, however, many students experience MD without a formal disability diagnosis (Szűcs & Goswami, 2013). In this study, similar to other research teams (Branum-Martin et al., 2012; Bryant et al., 2016; Fuchs et al., 2014), 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 without formal identification of a disability.

Difficulties with Mathematics

Across mathematics content, students experiencing MD frequently demonstrate lower mathematics performance across grade levels (Koponen et al., 2018; Martin et al., 2013) and lower proficiency levels on mathematics tests than students without MD (Cowan & Powell, 2014). For example, 70% of children who score below the 10th percentile in mathematics at the end of kindergarten receive an identification of MD by fifth grade (Morgan et al., 2009), and over 95% of students experiencing MD in fifth grade continue to demonstrate performance below the 25th percentile in high school (Shalev et al., 2005), indicating the persistence of MD.

Students experiencing MD may have difficulty with tasks involving counting (Stock et al., 2010), arithmetic (Tolar et al., 2016), whole-number computation (Raghubar et al., 2009), comparison (De Smedt & Gilmore, 2011), rational-number understanding (Fuchs et al., 2013), algebra (O'Shea et al., 2017), and mathematics vocabulary (Powell, Berry, et al., 2020). Word-problem solving proves especially challenging for students with MD (Krawec et al., 2012), who are at greater risk for school failure (Wei et al., 2013).

Although many students describe word problems as difficult (Jitendra et al., 2007), students experiencing MD demonstrate significantly lower scores on word-problem measures than students without MD (Lai et al., 2015; Peake et al., 2015). Students experiencing MD also make significantly more word-problems errors (Kingsdorf & Krawec, 2014). Beginning in the elementary grades, students experiencing MD struggle to solve addition and subtraction word problems because of the multiple steps required to develop a solution (Tolar et al., 2016). For example, word problems often require students to read a key and number a graph, understand the problem situation, build the situation model, determine the needed operation(s) for solving the problem, interpret and evaluate the problem, solve the problem correctly, and add a label corresponding to the number answer (Verschaffel et al., 2000). Without explicit instruction on how to set up and solve word problems, students experiencing MD exhibit frustration as they attempt to solve word problems without any clear procedure or steps. Thus, many students experiencing MD attend to superficial cues in the word problem and add or subtract without interpreting or considering a mathematical model. Frequently, students experiencing MD select the incorrect operation(s) for solving the word problem, misuse irrelevant information, and fail to develop a mental model based on the text description (Kingsdorf & Krawec, 2014; van Lieshout & Xenidou-Dervou, 2018; Wang et al., 2016).

Word-Problem Intervention Focused on Schemas

Word-problem intervention specifically focused on schemas has proven beneficial for students with MD (Cook et al., 2019; Fuchs et al., 2014; Jitendra et al., 2015). As defined by Marshall (1995), a *schema* allows for organization of an experience that can be recognized in other similar experiences. In word-problem solving, we use the term *schema* to refer to the conceptual word-problem structure or word-problem type (e.g., this problem is about parts and a total; that problem starts with an amount, then something happens to change the starting amount to a new end amount), and this schema knowledge can be used time after time to solve word problems with the same schema. 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). Starting in Grade 3, students solve word problems with the Equal Groups and Comparison schemas (Griffin et al., 2018; Xin & Zhang, 2009).

With the Total schema, also referred to as Combine (García et al., 2006), Group (Jitendra et al., 2007), or Part-Part-Whole (Peltier et al., 2020), students have parts that are put together for a total. For example, "A cat and dog take 7 naps during the day. If the dog takes 2 naps, how many naps does the cat take?" With the Difference schema, also named the Compare schema (Gvozdic & Sander, 2020; Jitendra et al., 2013), students compare two amounts for a difference. As an example, "A cat weighs 9 pounds and a dog weighs 45 pounds. How much more does the dog weigh than the cat?" With the Change schema, students have a start amount that increases or decreases to a new result. Change problems with an increase also may be named Join problems, and Change problems with a decrease may be referred to as Separate problems (Carpenter et al., 1981; van de Walle et al., 2019). A Change example with a decrease is, "The dog had 12 squeaky toys then the cat hid some. Now, the dog has 9 squeaky toys. How many toys did the cat hide?” In the Equal Groups schema, students multiply a quantity (i.e., groups) times a unit rate or size for a product. Equal Groups problems also may be named Scalar or Array (Agostino et al., 2010; Alghamdi et al., 2020). For example, “The dog has 3 dog beds with 3 dog bones in each bed. How many dog bones does the dog have?” With the Comparison schema, students multiply a set a number of times for a product. Xin (2008) referred to these as Multiplicative Comparison problems. As an example, “The dog jumped 2 feet. The cat jumped twice as high as the dog. How high did the cat jump?”

Several researchers have conducted meta-analyses and syntheses to understand the impact of schema-focused word-problem interventions for students experiencing MD. Jitendra et al. (2015) determined that schema-focused interventions, examined within 14 group design studies, led to improved word-problem outcomes. They calculated average effect sizes ranging from 1.27 (95% CI [0.93, 1.42]) to 1.29 (95% CI [0.86, 1.72]). In a meta-analysis of 21 studies primarily focused on students experiencing MD, Peltier and Vannest (2017) calculated an effect size of 1.57 (95% CI [1.52, 1.61]) favoring students who participated in schema-focused instruction. Lein et al. (2020) calculated smaller effect sizes when they analyzed 18 schema-focused interventions. For interventions in which students only learned about the schemas, they identified an effect size of 0.40 (95% CI [0.23, 0.58]). For interventions in which students learned the schemas and how to transfer schema knowledge, Lein et al. (2020) calculated an effect size of 1.06 (95% CI [0.88, 1.24]). Finally, in a review of schema-focused intervention for students with an identified learning disability, Cook et al. (2020) noted an effect size from one high-quality group design study of 1.69, with an average effect from five high-quality single-case designs of 0.87 (95% CI [0.67, 1.00]). This collection of research demonstrates the positive impact of schema-focused intervention for students experiencing MD.

At the individual study level, our research team (Powell et al., 2021) investigated the efficacy of a 16-week schema-focused word-problem intervention with embedded pre-algebraic reasoning instruction delivered individually to third-grade students experiencing MD. We screened all students for MD using a word-problem measure and identified students who scored at or below the 25th percentile as experiencing MD, a common cut-off score in research related to MD (Nelson & Powell, 2018). Results indicated that intervention students with MD significantly and substantially outperformed students in the business-as-usual (BaU) comparison (ES = 2.66) on a proximal measure of word-problem solving with 26 questions. Although results from previous studies and our recent work are promising, additional research is warranted to determine how individual interventions replicate within small-group settings, which are more reflective of supplemental, targeted, or Tier-2 mathematics intervention and offer greater feasibility for educators and remediation support teams. The purpose of this pilot study was to determine whether implementation of our schema word-problem intervention with small groups of third-grade students experiencing MD led to improved word-problem performance.

Mathematics Intervention in Research and in Practice

In much of the research conducted at the elementary level for students experiencing MD, researchers provide support to individual students through one-to-one tutoring (Alghamdi et al., 2020; Burns, 2005; Dennis, 2015; Fuchs et al., 2008; Fuchs, Geary, et al., 2013; Fuchs et al., in press; Powell et al., 2015; Xin & Zhang, 2009). In other studies, researchers have implemented mathematics interventions in small groups (Bryant et al., 2016; Clarke et al., 2014; Codding et al., 2016; Doabler et al., 2019; Flores et al., 2014; Fuchs et al., 2013; Jitendra et al., 2013; Peltier et al., 2020; Swanson et al., 2014). For example, Bryant et al. (2011) tutored first-grade students experiencing MD in groups of 3 to 5, whereas Fuchs et al. (2013) tutored fourth-grade students in groups of 3. And in other cases, researchers have determined the efficacy of intervention provided to classrooms of students with data analysis focused on students experiencing MD within those classrooms (Griffin et al., 2018).

Both Clarke et al. (2017) and Doabler et al. (2019) studied mathematics intervention effects for kindergarten students experiencing MD and noted no differences on several outcome measures regardless of whether students received tutoring in groups of 2 or groups of 5. They did, however, note that students tutored in groups of 2 had a greater number of opportunities to respond, whereas students tutored in groups of 5 had more opportunities to practice mathematics with their peers. What researchers have not examined is whether a mathematics intervention designed for individual implementation can be replicated in small groups. Such an investigation is critical for helping educators understand whether an efficacious intervention developed and tested in research settings can be replicated in practice (i.e., real-world settings).

In a small-group setting, a greater number of students experiencing MD can receive targeted support. Small-group intervention also offers an opportunity for educators to identify students who may need more intensive and individualized assistance (Barrett & VanDerHeyden, 2020). Providing intervention in small groups or at the classroom level is significantly cheaper than implementing an intervention individually to a student experiencing MD. Small-group settings also allow schools to provide intervention to as many students as possible (Clarke et al., 2017). Moreover, developing interventions for use in small-group settings aligns well with multi-tiered systems of supports (MTSS), which are implemented in schools across the U.S.. In a MTSS framework, at-risk students receive small-group tutoring (i.e., Tier 2; Fuchs et al., 2010) in a format that allows educators to determine whether more individualized tutoring is warranted.

Purpose and Research Questions

The purpose of this pilot study was to determine whether positive results (ES = 2.66 on 26 word problems) from our individually-delivered word-problem intervention (Powell et al., 2021) translated to similar word-problem outcomes for students with MD when implemented in a small-group setting. This pilot study was implemented with small groups of 3 to 4 third-grade students experiencing MD rather than as an individual intervention with the interventionist working one-on-one with the student. Our research questions were as follows: (1) What is the impact of a small-group word-problem intervention (Pirate Math Equation Quest: PMEQ) on outcomes related to word-problem solving for third-grade students experiencing MD? (2) Is the effect of PMEQ different when implemented in small groups versus individually?

Method

Context and Setting

After receiving approval from our university's Institutional Review Board and from our local school district to conduct research in public schools, we recruited elementary schools from a large urban school district in the Southwest of the United States. This public school district serves over 80,000 students. On average, the district reports 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 identify as dual-language learners, 52.4% qualify as economically disadvantaged, and 12.1% receive special education services. During the 2018-2019 school year, 52% of Grade 3 students in the school district met grade level standards on the state-level mathematics test administered at the end of the Grade 3 year.

Participants

During the 2018-2019 school year, we recruited 19 third-grade educators from 4 elementary schools. From these 19 classrooms, we screened 304 third-grade students. We screened all students using two word-problem performance measures: *Single-Digit Word Problems* (Jordan & Hanich, 2000) and *Texas Word Problems-Brief* (Powell & Berry, 2015). We selected both measures to screen for mathematics difficulty (MD) in the area of word problems because the primary focus of the intervention was word-problem solving. For study eligibility, we identified students who answered 7 or fewer items correctly (out of 14) on *Single-Digit Word Problems* and/or those who answered 4 or fewer items correctly (out of 8) on *Texas Word Problems-Brief* as experiencing MD. These two cut-off scores of 7 and 4 represented performance at or below the 25th percentile based on cut-off scores from Powell et al. (2020). The 25th percentile is 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 and completion of the pretest battery, we identified 131 third-grade students with word-problem MD. Of the 131 identified students with MD, 22 were deemed ineligible for participation in the intervention for the following reasons: behavior challenges identified by interventionist during screening (n = 8), limited English proficiency (n = 6), incomplete screening assessment (n = 2), withdrawal from screening by educator (n = 1), identification of intellectual disability by the educator (n = 1), numerous special education pull out time requirements (n = 3), and no consent (n = 1). Of the remaining 109 students with MD in the 19 classrooms, we randomly selected 4 students in each of the 19 classrooms for participation in the present study. We made the decision to have only one small group of 4 students per classroom because our tutoring team could only tutor 14 groups, and we did not want group size to exceed 4 students. This decision eliminated 33 students experiencing MD from being eligible for tutoring. Thus, we included 76 students from 19 classrooms (4 in each classroom) in the present study. Table 1 displays the demographics for the 76 students included in the present study.

Random Assignment

The 76 students represented 19 classrooms with 4 students with MD from each classroom. We randomly assigned, blocking on school, the classrooms to one of two conditions: Pirate Math Equation Quest (PMEQ) word-problem intervention (*n* = 14 classrooms with *n* = 56 students) or business-as-usual (BaU) comparison (*n* = 5 classrooms with *n* = 20 students). We included over two-thirds of the classrooms in the PMEQ intervention condition for two reasons. First, based on our prior research (Powell et al., 2021), we learned that students who participated in word-problem intervention demonstrated significant gains over students in a BaU comparison, with an effect size of 2.66. Therefore, we understood the efficacy of the individually-administered intervention and the limited growth on word-problem performance for students in a BaU. Second, we wanted to maximize the number of groups and students receiving the small-group word-problem intervention to understand the effects of such grouping.

Random assignment occurred as follows. We assigned random numbers to each classroom, and then sorted, by school, the random numbers seven times. In the school with three classrooms, we assigned the first two classrooms to PMEQ and the third classroom to BaU. In the school with six classrooms, we assigned the first two classrooms to PMEQ and the third classroom to BaU, and then we repeated the same pattern for the fourth, fifth, and sixth classrooms. In the two schools with five classrooms, we assigned the first two classrooms to PMEQ, the third classroom to BaU, and next two classrooms to PMEQ.

Overall, 8 students (10.5% of the 76 randomized students) did not complete the intervention because they (a) left the participating school prior to treatment's end (*n* = 3), (b) were discontinued from intervention due to disruptive behavior (*n* = 1), (c) had a parent who opted out of the study (*n* = 2), (d) were truant (*n* = 1), and (e) had conflicts with special education schedule (*n* = 1). Attrition rates varied across treatment conditions. In BaU, all students completed the posttest battery, while 8 students in the PMEQ intervention did not complete posttesting for the aforementioned reasons. The 8 PMEQ students who did not complete posttesting came from 8 different classrooms from the 4 different schools. These students left intervention after session 1, 6, 9, 20, 23, 27, 31, or 38.

General Education Instruction

All 76 students experiencing MD participated in regular mathematics instruction provided by their general educator. In the district, educators primarily used the *GO Math!* and *Bridges in Mathematics* curricula to guide mathematics instruction. Students in the PMEQ condition also received our supplemental intervention about word-problem solving in small groups of 3 to 4. 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.

Alterations from Individual to Small-Group Word-Problem Implementation

In Powell et al. (2021), we determined the efficacy of a 16-week word-problem intervention designed for third-grade students experiencing MD. We named this intervention Pirate Math Equation Quest (PMEQ). Students learned to solve word problems following Pirate Math strategies (Fuchs et al., 2008) with an embedded pre-algebraic reasoning component called Equation Quest. Interventionists delivered the PMEQ intervention one-to-one in 30-min sessions delivered 3 times per week. Five activities occurred each session with a focus on the three additive word-problem schemas: Total, Difference, and Change.

In the present study, we made several changes. Table 2 provides an outline of the individual PMEQ intervention versus the small-group PMEQ intervention. First, to better align the intervention with the school district's calendar, the small-group intervention followed a 13-week intervention duration rather than the longer 16 weeks of tutoring provided in the individual intervention. Second, the small-group PMEQ included a focus on all three additive word-problem schemas (i.e., Total, Difference, and Change) as well as the Equal Groups multiplicative schema. We included the Equal Groups schema because of expectations within the state on the high-stakes exam for students to solve word problems related to all four schemas. One-to-one PMEQ did not include Equal Groups problems. Third, we redesigned the five lesson activities for each intervention session using a round-robin format, with interventionists asking students to take turns answering questions in a circle around the table.

For the students assigned to the PMEQ condition, interventionists conducted sessions three times per week for 13 weeks (i.e., 39 completed sessions) for 30 min a session. Rather than working one-on-one with students as in the individual intervention, interventionists worked with small groups of 3 to 4 students in a quiet place outside of the classroom (e.g., school library, conference room, extra classroom). PMEQ students participated in five activities during each session. In the following sections, we describe each of the five activities, including the adjustments made to the individual intervention to support learning in small-group settings.

Math Fact Flashcards

Lessons 1-30, interventionists displayed addition and subtraction flashcards (addends 0 to 9; minuends 0 to 18; and subtrahends 0 to 9). For Lessons 31-39, interventionists displayed multiplication and division flashcards (factors 0 to 11; dividends 1 to 121; and divisors 1 to 11). To ensure all students in the group received the same number of learning opportunities, interventionists used a round robin and instructed students to take turns answering questions in a circle. After setting the timer, students answered as many flashcards as they could in 1 min. After 1 min of the round robin, interventionists and students counted the number of flashcards answered correctly. Prior to starting a 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 approximately 2 to 5 min each session, interventionists provided instruction on solving equations and the meaning of the equal sign. Students learned the equal sign acts as a balance between two sides of an equation and does not solely signal a calculation. To interpret 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). Interventionists posed questions and elicited responses from students using a round robin format to ensure even participation among all group members.

Buccaneer Problems

The third activity for each session consisted of interventionist-led schema instruction through a series of three Buccaneer Problems. 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 utilized a round robin method to ask questions and encourage students to model how they solved each Buccaneer Problem.

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 students. Each square was labeled with one word-problem type. Interventionists reminded students to sort the word-problem cards and to not solve any of the word problems. Shipshape Sorting followed a timed round robin format in which students took turns answering questions in a circle. Interventionists set the timer for 1 min and read the first word-problem card aloud before handing the card to the first student in the round robin. After 1 min, interventionists provided immediate, corrective feedback by reviewing at least three of the word-problem cards with students.

Jolly Roger Review

The final component of each session, the Jolly Roger Review, served as an independent practice activity and included a brief, timed paper-and-pencil review of the session content. Students worked individually for 1 min to answer math facts, solve computation problems, or write appropriate equations for the four word-problem schemas. Then, students worked individually for 2 min to solve a word problem using the schema steps taught during the Buccaneer Problems. At the end of the 3 min, interventionists briefly reviewed the correct responses and provided feedback to the small group.

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 from their educators. Classroom word-problem instruction for students in the BaU condition incorporated general mnemonic devices (e.g., CUBES: Circle the numbers, Underline the question, Box math action words, Eliminate, Solve then; UPS Check: Understand, Plan, Solve, Check), key word clues (e.g., *altogether* means add), and practice in applying problem-solution rules, as self-reported by participating educators. Notably, no classroom educators utilized schema instruction.

Interventionists

We recruited 5 interventionists to conduct the pretesting, tutoring, and posttesting. All interventionists were pursuing or had obtained a Master's or doctoral degree in an education-related field. All interventionists were female (n = 5), with 60% (n = 3) identifying as Caucasian, 20% percent as Hispanic (n = 1), and 20% as Asian American (n = 1). Throughout the year, interventionists participated in trainings to ensure strong preparation for all aspects of the intervention. In late August, interventionists participated in three, 1.5-hr pretesting trainings. In early October, the team participated in a 1.5-hr tutoring training about the content of the intervention and Total problems. Three subsequent 1.5-hr tutoring trainings followed in late October to introduce Difference problems, late November to introduce Change problems, and early January to introduce Equal Groups problems. Lastly, interventionists participated in one, 1.5-hr posttesting training meeting in late January.

Fidelity of Implementation

We collected fidelity of implementation in several ways. First, for pretesting and posttesting, interventionists recorded all testing sessions. We randomly selected >20% of audio recordings for analysis, evenly distributed across interventionists, and measured fidelity to testing procedures against detailed fidelity checklists. We measured pretesting fidelity at 97.9% (*SD* = 3.7%) and posttesting fidelity at 98.7% (*SD* = 2.8%).

Second, we measured fidelity of implementation of the interventions. The Project Manager developed a unique fidelity checklist for each of the 39 sessions and conducted in-person fidelity observations once every two weeks for every interventionist. During the fidelity observation, the Project Manager scored the interventionist across each of the five intervention components: *Math Fact Flashcards, Equation Quest, Buccaneer Problems, Shipshape Sorting,* and *Jolly Roger Review*. Each fidelity checklist included 50 to 150 items, depending on the length and difficulty of the session content. Example checklist items included: *Interventionist explains that students will have 1 min to solve as many problems as they can as a group, using a round robin* or *Interventionist and students read problem A, following the RUN guide.* Directly after the observation, the Project Manager scored the fidelity checklist as a percentage out of 100. They provided positive feedback to the interventionist, reviewed the missed checklist items, and offered ideas for improving performance. For copies of the 39 fidelity checklists, please make a request to the project team.

We also measured fidelity of intervention implementation through analysis of audio-recorded sessions. We audio-recorded every intervention session and selected >20% of audio-recorded sessions for analysis, evenly distributed across interventionists. Fidelity averaged 97.7% (*SD* = 2.2%) for in-person supervisory observations and 97.8% (*SD* = 5.5%) for audio-recorded intervention sessions.

Measures

Screening Measures

We used *Single-Digit Word Problems* (Jordan & Hanich, 2000; Powell & Berry, 2015) as one measure for identifying students with MD, and we administered this assessment in a whole-class 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. Interventionists read each word problem aloud and could re-read each problem up to one time upon student request. Interventionists provided approximately 1 min for students to solve each problem, but we did not time the test administration. We scored *Single-Digit Word Problems* as the number of correct responses (maximum = 14). We calculated Cronbach's α as .87.

During the whole-class screening, 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. For each problem, interventionists read the problem aloud and provided approximately 1 to 1.5 min for students to solve the problem and write an answer. Interventionists could re-read each problem up to one time upon student request. We did not time the test administration. We scored the measure as the number of correct numerical and label responses for a maximum score of 16. Cronbach's α was .79.

Pre- and Posttest Measures

Interventionists conducted two, 45-min pretesting sessions with groups of four students or fewer. In the first pretesting session, interventionists 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. Interventionists read each problem aloud and provided students time (approximately 1 to 1.5 min) to solve the problem and write an answer. Interventionists could re-read each problem up to one time upon student request. We did not time the test administration. We scored this measure as the number of correct numerical and label responses, with a maximum score of 18.

We also administered *State of Texas Assessments of Academic Readiness (STAAR)-Part 1* (Berry & Powell, 2018) in the first pretesting session. With this measure, we intended to capture students' understanding of word problems presented on the Texas standardized test, called the State of Texas Assessments of Academic Readiness. To develop *State of Texas Assessments of Academic Readiness (STAAR)-Part 1,* we visited the Texas Education Agency website and reviewed mathematics released items from the 2013, 2014, 2016, and 2017 STAAR tests (i.e., 2015 was not released). From these assessments, we selected eight word problems (four administered during *Part 1*; four during *Part 2*). For *Part 1,* students solved four word problems: one Equal Groups problem involving area, one double Change problem, one Total problem, and one Difference problem. Three of the four problems in *Part 1* included a table, figure, or model. Interventionists read each problem aloud and provided students time (approximately 2 to 3 min) to solve the problem and write an answer. Interventionists could re-read each problem up to one time upon student request. We did not time the administration of this measure. We scored this measure as the number of correct numerical and label responses, with a maximum score of 8.

In the second pretesting session, interventionists 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. Interventionists administered this identically to *Texas Word Problems-Part 1.* The maximum score was 18.

We also administered *State of Texas Assessments of Academic Readiness (STAAR)-Part 2* (Berry & Powell, 2018) in the second pretesting session. For *Part 2*, students solved four word problems: one Equal Groups problem, one Total problem involving perimeter, one Difference problem, and one double Change problem. One of the four problems in *Part 2* included a figure. Interventionists followed identical administration procedures as in *Part 1*.

For posttesting, interventionists administered the same assessments as in pretesting, following identical procedures with small groups of 4 students. Although we administered other measures across the whole-class screening and two small-group pre- and posttesting sessions, we only discuss the word-problem measures used in the present study. The other measures administered at screening included a test of equation solving and a test of mathematics vocabulary. We did not administer either of these measures at posttest. At pretest and posttest, we included tests of single-digit addition, single-digit subtraction, double-digit addition, and double-digit subtraction.

In terms of validity of these measures, a separate research team at another university developed *Single-Digit Word Problems*. Research teams have used this measure as a screener and outcome measures in a number of studies (Driver & Powell, 2017; Fuchs et al., 2014; Fuchs et al., 2021; Hanich et al., 2001; Jordan et al., 2007; Vukovic et al., 2013). As displayed in Table 3, *Single-Digit Word Problems* demonstrated significant correlations with the other five word-problem measures used in this study. Our research team developed each of the *Texas Word Problems* measures. We developed these measures by collecting word problems representing each of the schemas and positions of the unknown within the word problem, but we did not conduct a formal analysis of the content validity. Correlations with the *Texas Word Problems* screener showed moderate and significant correlations with the *Texas Word Problems* pretest and posttest as well as the *STAAR.* We noted a similar trend for *Texas Word Problems* pretest and posttest when compared to the *STAAR* pretest and posttest. Pearson developed items from the 2013 and 2014 STAAR, and Educational Testing Service developed items from the 2016 and 2017 STAAR. The Texas Education Agency presented content validity evidence for the Grade 3 mathematics version of the *STAAR.* We noted moderate to strong correlations between the *STAAR* and *Texas Word Problems*.

Scoring

Two interventionists 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 compared the discrepancies between the two databases across each outcome measure and rectified any inconsistencies to reflect the original response. Two interventionists 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 99.8% for pretesting and posttesting.

For our analysis, we created a composite score for proximal word-problem performance by combining *Texas Word Problems-Part 1* and *Texas Word Problems-Part 2.* We calculated Cronbach's α at .92. We created a composite score for distal word-problem performance by combining *State of Texas Assessments of Academic Readiness (STAAR)-Part 1* and *State of Texas Assessments of Academic Readiness (STAAR)-Part 2*. Cronbach's α was .87.

Procedure

During the first week of September, we administered a whole-class screening in one, 55-min session. Identification of students with MD occurred shortly thereafter, with four weeks of small-group pretesting for eligible students during the last two weeks of September and the first two weeks of October. During the third week of October, approximately 4 to 6 days after the completion of pretesting, intervention began and occurred three times per week for 13 weeks, concluding the last week of January. Approximately 4 to 6 days after the last intervention session, posttesting occurred in two, 45-min small group sessions with four students or fewer. We administered all of posttesting during the first week of February. We pre- and posttested all BaU students in the same time frame as the intervention students.

Data Analysis

To estimate the impact of Pirate Math Equation Quest (PMEQ) implemented in small groups on outcomes related to word-problem solving, we fit multilevel models. Because randomization happened at the classroom level, the impact of PMEQ was estimated at the class-level with experimental condition indicated by a dummy code (1 = PMEQ, 0 = BaU). In the unconditional model, 66% of the variance in proximal word problem and 57% of variance in the distal word problem measure was associated with classroom. An additional 9% of the variance in the distal word problem measure was at the school level. On the proximal word problem measure, variance at the school level was zero. For both the distal and proximal outcomes, the interventionist-related variance was zero. Accordingly, we modeled data as two-level for the proximal word problem and three-level for the distal word problem and estimated main treatment effects at the classroom level. Pretest scores were group-mean centered (Enders & Tofighi, 2007) and used as a student-level covariate. Additionally, classroom-level mean scores at pretest were grand-mean centered and included as a level-2 covariate to minimize class-level variability and improve the power of effect estimates (Hox et al., 2017).

The reduced-form equation for the three-level model was:

$$Distal\ word\ problem_{ijk} = \gamma_{000} + \gamma_{100}\,(Pretest)_{ijk} + \gamma_{010}(Pretest)_{jk} + \gamma_{020}PMEQ_{jk} + e_{ijk} + r_{jk} + u_k$$

Here, *i* represents students, *j* represents classrooms, and *k* represents schools. Parameter γ_{000} represents the mean student outcome across all classrooms and schools; γ_{100} represents student-level pretest scores centered around the classroom mean; γ_{010} is the classroom-level pretest aggregate centered around the grand mean; $PMEQ_{jk}$ is class-level dummy-coded variable representing assignment to the PMEQ intervention or BaU; residuals e_{ijk}, r_{jk}, and u_k are Level 1, 2, and 3 random effects, respectively. We used Hedges' (2011) equations for cluster-randomized two-level and three-level models to estimate effect sizes at the student level.

Results

Baseline Equivalence

We tested for baseline equivalence between PMEQ and BaU conditions on the two pretest composites of interest (What Works Clearinghouse, 2017). The two groups differed at pretest on mean classroom-level scores, suggesting nonequivalence prior to onset of treatment. More specifically, students in the PMEQ condition had higher pretest scores on the proximal word-problem measure (g = 0.77, 95% CI [0.03, 1.51]) and distal word-problem measure (g = 0.69, 95% CI [0.02, 1.36]) than students in the BaU condition. This represents considerable non-equivalence between treatment conditions at baseline. We included pretest scores as student-level (group-mean centered) and classroom-level (grand mean centered) covariates.

Impact of PMEQ on Word-Problem Outcomes

We conducted a post-hoc power analysis with alpha of 0.05. This post-hoc analyses demonstrated we had substantial power to detect differences between the two conditions. Table 4 summarizes observed pretest and posttest means and standard deviations for each measure at student and classroom levels. For our first research question, we investigated the impact of PMEQ on outcomes related to word-problem solving. Students in classrooms assigned to the PMEQ significantly outperformed students in BaU classrooms on proximal word problems. As shown in Table 5, after adjusting for differences at pretest, students in classes assigned to the PMEQ scored 5.67 points higher at posttest than students in BaU classes ($\beta = 5.67$, $p = .04$). Hedges' g was 0.76, 95% CI [0.05, 1.46]. The estimate of PMEQ impact on the distal word-problem measure was not statistically significant ($\beta = 1.85$, $p =.19$, $g = 0.51$, 95% CI [-0.24, 1.25]).

Individual Versus Small-Group Implementation

With our second research question, we explored the effect of PMEQ when implemented in small groups versus individually. To compare the effect of PMEQ when implemented in small groups versus individually, we reran the model for the proximal word-problem outcome ignoring clustering and calculated the effect size as the covariate-adjusted mean difference divided by the unadjusted pooled within-group SD (What Works Clearinghouse, 2020). As shown in Table 6, when implemented in small groups, ignoring the clustering at the classroom level, the effect of PMEQ was 0.83, 95% CI [0.57, 1.10]. We reanalyzed the data set from Powell et al. (2021) to look at student performance on the same 18 word problems as the proximal word-problem outcome used in the present study. The effect of PMEQ implemented individually was 1.99, 95% CI [1.82, 2.16].

Discussion

We conducted this pilot study to determine the potential impact of using a word-problem intervention designed for individual implementation with small groups of students. Our primary reason for conducting this research was to help district-, school-, and classroom-level administrators and educators understand the degree to which a one-to-one intervention could be used in small groups. As most schools provide supplemental, Tier-2 mathematics support in small groups (Fuchs & Fuchs, 2007; Schumacher et al., 2017), such knowledge proves useful to educators by increasing the number of mathematics interventions available for use. This information is especially important in the area of mathematics because the number of available efficacious mathematics interventions is far less than the number available for reading (National Center on Intensive Intervention, 2020).

With our first research question, we explored the impact of PMEQ on word-problem outcomes. On the proximal word-problem measure, we determined that, after participation in the 39 sessions, PMEQ students demonstrated improved proximal word-problem performance compared to students in the BaU, with an effect size of 0.76. This result corroborates prior research about word-problem intervention focused on schemas and implemented individually (Alghamdi et al., 2020; Flores et al., 2016; Fuchs et al., in press; Powell & Fuchs, 2010; Xin & Zhang, 2009) or in small groups (Fuchs, Powell, et al., 2014; Jitendra et al., 2013; Morin et al., 2017; Peltier et al., 2020). Use of PMEQ in small groups (of 3 to 4 students) led to improved proximal word-problem performance; students who learned about word-problem solving from the intervention demonstrated a higher rate of word-problem growth than students who did not participate in the intervention. Results from this pilot study suggest the benefit of PMEQ when implemented in small groups of students with MD. Researchers and educators also may use our findings to explore whether other mathematics interventions designed for use in individual settings effectively translate to small-group settings with minimal adaptations. Our primary adaptation included altering the way the interventionist interacted with the students, moving from individual questioning to round-robin participation and transitioning from questions answered by individual students to questions answered chorally by the small group. Understanding the degree to which individual interventions can be used successfully in small groups offers potential for increasing the number of available mathematics interventions for use within MTSS.

On the distal word-problem measure comprised of high-stakes mathematics test released items, PMEQ students demonstrated greater growth than BaU students; however, the difference was not significant (ES = 0.51). We hypothesize that, with a larger sample size, we may have detected a significant difference between the two conditions. Our result, however, mirrors prior research in which author teams noted significant gains on proximal measures but not distal measures (Fuchs, Powell, et al., 2014; Fuchs et al., 2008; Jitendra et al., 2013; Powell, Fuchs, et al., 2015). Future research should investigate the degree to which PMEQ strategies transfer to distal word-problem tasks and examine why distal problems cause difficulty for students experiencing MD. Future research should consider providing more opportunities within the intervention sessions for students to practice problems that look similar and use similar vocabulary to distal high-stakes items. With such practice, however, the distal items would be considered more as proximal items. We would suggest for researchers to collect data from school districts about student performance on high-stakes, yearly-administered mathematics tests to determine whether participation in efficacious interventions transfer to school-administered tests featuring word problems.

With our second research question, we examined whether response to PMEQ was different when implemented in small groups versus the original individual implementation of Powell et al. (2021). In the original implementation, we calculated an ES of 2.66 on a word-problem composite comprised of 26 items. In the present study, we only administered 18 of the same 26 items at pre- and posttest in the small-group study; therefore, we reran an analysis of the dataset from Powell et al. (2021) to compare individual implementation versus small-group implementation on the word-problem composite with 18 items. While both versions (i.e., small group and individual) of PMEQ led to improved proximal word-problem outcomes, the effect for PMEQ over BaU was 0.83 when implemented in small groups versus 1.99 when implemented individually in Powell et al. (2021). Given the financial and time constraints of many schools (Barrett & VanDerHeyden, 2020), we suggest using PMEQ in small groups whenever a group of students experiencing MD requires supplemental and targeted word-problem support. Our suggestion is similar to that of Clarke et al. (2017), who debated the impact of mathematics intervention delivered in groups of 2 or 5; even though students in the groups of 5 received fewer practice opportunities, the authors suggested that educators tutor in groups of 5 to meet the needs of as many students as possible.

Limitations

Before concluding, we note several limitations to this pilot study. First and foremost, this was a pilot study. We had a small sample size, and we oversampled groups of students into the PMEQ condition. Future research should recruit more schools and classrooms, across multiple cohorts and sites, to understand the true impact of PMEQ implemented in group settings. Future research also may want to investigate the size of groups, similar to research conducted by Clarke et al. (2017) and Doabler et al. (2019).

Second, in order to compare PMEQ implemented in small groups versus individually, we did a reanalysis of the data from Powell et al. (2021) to understand the effect of the individually-administered PMEQ. In Powell et al. (2021), the authors administered three tests of proximal word problem solving: *Texas Word Problems-Brief, Texas Word Problems-Part 1,* and *Texas Word Problems-Part 2*. These three tests were summed for a composite word-problem outcome with 26 word problems. In the present study, we only administered *Texas Word Problems-Part 1* and *Part 2* at posttest with a total of 18 word problems. Therefore, to compare the small-group implementation of PMEQ to the individual implementation of PMEQ, we reran the analysis of Powell et al. (2021) with the same 18 problems as used in the present study. This accounts for a difference in effect sizes from Powell et al. (2021) of 2.66 favoring one-to-one PMEQ over the BaU to the effect size in this study of 1.99 favoring one-to-one PMEQ over the BaU. In subsequent comparisons of the same program implemented in different ways, we would administer the exact same test battery to students.

Third, our comparison of small-group PMEQ was not a direct comparison to one-to-one PMEQ. In this pilot study, students participated in 39 sessions of PMEQ, whereas students in one-to-one PMEQ participated in a minimum of 45 sessions. Furthermore, students in small-group PMEQ practiced four schemas (Total, Difference, Change, Equal Groups) versus three schemas from individually-administered PMEQ (Total, Difference, Change). We emphasize that we did not conduct a direct comparison of PMEQ implemented individually to PMEQ implemented in small groups. To accurately compare the impact of different settings, future research should recruit schools and classrooms and randomly assign students with MD to receive PMEQ in small groups or individually. This design also should involve collecting more information about the interaction between the interventionist and student or group of students (Doabler et al., 2019) to understand the mechanisms driving any differences in student-level performance.

Fourth, it is difficult for us to determine whether the small-group intervention led to improved word-problem outcomes. Another factor could be that small-group PMEQ students demonstrated improved outcomes because they spent more time practicing mathematics with an interventionist. In the individual intervention study (Powell et al., 2021), we ran two competing word-problem interventions in which one condition (Pirate Math [PM]-alone) was a word-problem comparison to the intervention (PMEQ) used in the present study. In Powell et al. (2021), students in the competing word-problem intervention of PM-alone showed improved word-problem outcomes from pre- to posttest but PMEQ students showed higher gains than PM-alone students. Furthermore, with a sequential mediation model, Powell et al. (2021) demonstrated an advantage to using PMEQ over PM-alone. Because we compared PMEA to another word-problem intervention in the individual intervention study, we did not see the need to do the same comparison in the present study.

Next, our research team hired and trained the interventionists who implemented the intervention. We pulled students from their classrooms for tutoring. Future research should study the effects of PMEQ implemented in small groups by general educators, special educators, mathematics interventionists, or other support staff. Future research also should investigate the degree to which PMEQ can be implemented in the classroom during small-group instruction or within a workshop model and how PMEQ can be used within a school's MTSS framework to provide mathematics support to at-risk students (Schumacher et al., 2017).

Finally, and perhaps most notably, the PMEQ and BaU groups differed significantly at pretest. The small samples and classroom-level randomization may be responsible for the apparent failure of randomization. The allocation of matched pairs (classroom level) to conditions might have optimized balance at pretest. Our covariate-adjusted mean differences should be interpreted accordingly.

Conclusion

We determined students experiencing MD in third grade benefitted from participation in the PMEQ word-problem intervention when implemented by interventionists in groups of 3 to 4. While the effect size was lower for the small-group iteration of PMEQ when compared to the original, individual iteration of PMEQ (Powell et al., 2021), such findings have practical implications for educators. The number of students in U.S. schools who do not meet minimum levels of mathematics proficiency is around 60% at fourth grade (National Assessment of Educational Progress, 2019); therefore, a majority of students should have access to supplemental mathematics support in which educators use efficacious interventions. When students experiencing MD need additional help to solve mathematics word problems, educators should consider implementing PMEQ with small groups of students.

References

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

Agostino, A., Johnson, J., & Pascual-Leone, J. (2010). Executive functions underlying multiplicative reasoning: Problem type matters. *Journal of Experimental Child Psychology, 105*, 286–305. https://doi.org/10.1016/j.jecp.2009.09.006

Barrett, C. A., & VanDerHeyden, A. M. (2020). A cost-effectiveness analysis of classwide math intervention. *Journal of School Psychology, 80*, 54–65. https://doi.org/10.1016/j.jsp.2020.04.002

Berry, K. A., & Powell, S. R. (2018). *State of Texas Assessments of Academic Readiness (STAAR) - Parts 1 and 2.* Available from S. R. Powell, 1912 Speedway, D5300, Austin, TX 78712.

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

Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J, & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children, 78*(1), 7–23. https://doi.org/10.1177/001440291107800101

Burns, M. K. (2005). Using incremental rehearsal to increase fluency of single-digit multiplication facts with children identified as learning disabled in mathematics computation. *Education and Treatment of Children, 28*(3), 237–249.

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trend in Cognitive Sciences, 14*(12), 534–541. https://doi.org/10.1016/j.tics.2010.09.007

Carpenter, T. P., Hiebert, J., & Moser, J. M. (1981). Problem structure and first-grade children's initial solution processes for simple addition and subtraction problems. *Journal for Research in Mathematics Education, 12*(1), 27–39. https://doi.org/10.2307/748656

Clarke, B., Doabler, C. T., Cary, M. S., Kosty, D., Baker, S., Fien, H., & Smolkowski, K. (2014). Preliminary evaluation of a Tier 2 mathematics intervention for first-grade students: Using a theory of change to guide formative evaluation activities. *School Psychology Review, 43*(2), 160–177. https://doi.org/10.1080/02796015.2014.12087442

Clarke, B., Doabler, C. T., Kosty, D., Nelson, E. K., Smolkowski, K., Fien, H., & Turtura, J. (2017). Testing the efficacy of a kindergarten mathematics intervention by small group size. *AERA Open, 3*(2), 1–16. https://doi.org/10.1177/2332858417706899

Codding, R. S., VanDerHeyden, A. M., Martin, R. J., Desai, S., Allard, N., & Perrault, L. (2016). Manipulating treatment dose: Evaluating the frequency of a small group intervention targeting whole number operations. *Learning Disabilities Research and Practice, 31*(4), 208–220. https://doi.org/10.1111/ldrp.12120

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

Cowan, R., & Powell, D. (2014). The contributions of domain-general and numerical factors to third-grade arithmetic skills and mathematical learning disability. *Journal of Educational Psychology, 106*(1), 214–229. https://doi.org/10.1037/a0034097

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

De Smedt, B., & Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology, 108*, 278–292. https://doi.org/10.1016/j.jecp.2010.09.003

Devine, A., Hill, F., Carey, E., & Szűcs, D. (2018). Cognitive and emotional math problems largely dissociate: Prevalence of developmental dyscalculia and mathematics anxiety. *Journal of Educational Psychology, 110*(3), 431–444. https://doi.org/10.1037/edu0000222

Doabler, C. T., Clarke, B., Kosty, D., Kurtz-Nelson, E., Fien, H., Smolkowski, K., & Baker, S. K. (2019). Examining the impact of group size on the treatment intensity of a Tier 2 mathematics intervention within a systematic framework of replication. *Journal of Learning Disabilities, 52*(2), 168–180. https://doi.org/10.1177/0022219418789376

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at any old issue. *Psychological Methods, 12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

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

Flores, M. M, Hinton, C., Strozier, S. D. (2014). Teaching subtraction and multiplication with regrouping using the concrete-representational-abstract sequence and strategic instruction model. *Learning Disabilities Research and Practice, 29*(2), 75–88.

Fuchs, L. S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children, 39*(5), 14–20. https://doi.org/10.1177/004005990703900503

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Schatschneider, C., Hamlett, C. L., DeSelms, J., Seethaler, P. M., Wilson, J., Craddock, C. F., Bryant, J. D., Luther, K., & Changas, P. (2013). Effects of first-grade number knowledge tutoring with contrasting forms of practice. *Journal of Educational Psychology, 105*(1), 58–77. https://doi.org/10.1037/a0030127

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., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., Jordan, N. C., Siegler, R., Gersten, R., & Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology, 105*(3), 683–700. https://doi.org/10.1037/a0032446

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

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

Fuchs, L. S., Zumeta, R. O., Schumacher, R. F., Powell, S. R., Seethaler, P. M., Hamlett, C. L., & Fuchs, D. (2010). The effects of schema-broadening instruction on second graders' word-problem performance and their ability to represent word problems with algebraic equations: A randomized control study. *The Elementary School Journal, 110*(4), 440–463. https://doi.org/10.1086/651191

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

Gvodic, K., & Sander, E. (2020). Learning to be an opportunistic word problem solver: Going beyond informal solving strategies. *ZDM Mathematics Education*, *52*, 111–123. https://doi.org/10.1007/s11858-019-01114-z

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

Hedges, L. V. (2011). Effect sizes in three-level cluster-randomized experiments. *Journal of Educational and Behavioral Statistics, 36*(3), 346–380. https://doi.org/10.3102/1076998610376617

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications.* Routledge.

Jitendra, A. K., Griffin, C. C., Deatline-Buchman, A., & Sczesniak, E. (2007). Mathematical word problem solving in third-grade classrooms. *The Journal of Educational Research, 100*(5), 283–302. https://doi.org/10.3200/joer.100.5.283-302

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

Jitendra, A. K., Rodriguez, M., Kanive, R., Huang, J.-P., Church, C., Corroy, K. A., &

Zaslofsky, A. (2013). Impact of small-group tutoring interventions on the mathematical problem solving and achievement of third-grade students with mathematics difficulties. *Learning Disability Quarterly, 36*(1), 21–35. https://doi.org/10.1177/0731948712457561

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

Kingsdorf, S., & Krawec, J. (2014). Error analysis of mathematical word problem solving across students with and without learning disabilities. *Learning Disabilities Research and Practice, 29*(2), 66–74. https://doi.org/10.1111/ldrp.12029

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

Koponen, T., Aro, M.m Poikkeus, A.-M., Niemi, P., Lerkkanen, M.-K., Ahonen, T., & Nurmi, J.-E. (2018). Comorbid fluency difficulties in reading and math: Longitudinal stability across early grades. *Exceptional Children, 84*(3), 298–311. https://doi.org/10.1177/0014402918756269

Krawec, J., Huang, J., Montague, M., Kressler, B., & Melia de Alba, A. (2012). The effects of cognitive strategy instruction on knowledge of math problem-solving processes of middle school students with learning disabilities. *Learning Disability Quarterly, 36*(2), 80–92. https://doi.org/10.1177/0731948712463368

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

Lein, A. E., Jitendra, A. K., & Harwell, M. R. (2020). Effectiveness of mathematical word problem solving interventions for students with learning disabilities and/or mathematics difficulties: A meta-analysis. *Journal of Educational Psychology, 112*(7), 1388–1408. https://doi.org/10.1037/edu000045

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

Marshall, S. P. (1995). *Schemas in problem solving.* Cambridge University Press.

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

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

Morin, L. L., Watson, S. M. R., Hester, P., & Raver, S. (2017). The use of a bar model drawing to teach word problem solving to students with mathematics difficulties. *Learning Disability Quarterly, 40*(2), 91–104. https://doi.org/10.1177/0731948717690116

Morsanyi, K., van Bers, B. M. C. W., McCormack, T., & McGourty, J. (2018). The prevalence of specific learning disorder in mathematics and comorbidity with other developmental disorders in primary school-age children. *British Journal of Psychology*, *109*, 917–940. https://doi.org/10.1111/bjop.12322

National Assessment of Educational Progress (2019). *The nation's report card.* Retrieved from https://nces.ed.gov/nationsreportcard/mathematics/

National Center on Intensive Intervention. (2020, November 28). *Intervention Tools Chart.* Retrieved from https://intensiveintervention.org/about-charts-resources

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

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

O'Shea, A., Booth, J. L., Barbieri, C., McGinn, K. M., Young, L. K., & Oyer, M. H. (2017). Algebra performance and motivation differences for students with learning disabilities and students of varying achievement levels. *Contemporary Educational Psychology, 50,* 80–96. https://doi.org/10.1016/j.cedpsych.2016.03.003

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

Peltier, C., Sinclair, T. E., Pulos, J. M., & Suk, A. (2020). Effects of schema-based instruction on immediate, generalized, and combined structured word problems. *The Journal of Special Education, 54*(2), 101–112. https://doi.org/10.1177/0022466919883397

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

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

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

Powell, S. R., Berry, K. A., & Tran, L. M. (2020). Performance differences on a measure of mathematics vocabulary for English Learners and non-English Learners with and without mathematics difficulty. *Reading and Writing Quarterly: Overcoming Learning Difficulties, 36*(2), 124–141. https://doi.org/10.1080/10573569.2019.1677538

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://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., 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., Namkung, J. M., & Lin, X. (in press). An investigation of using keywords to solve word problems. *The Elementary School Journal.*

Raghubar, K., Cirino, P., Barnes, M., Ewing-Cobbs, L., Fletcher, J., & Fuchs, L. (2009). Errors in multi-digit arithmetic and behavioral inattention in children with math difficulties. *Journal of Learning Disabilities, 42*(4), 356–371.

https://doi.org/10.1177/0022219409335211

Schumacher, R. F., Zumeta Edmonds, R., & Arden, S. V. (2017). Examining implementation of intensive intervention in mathematics. *Learning Disabilities Research and Practice, 32*(3), 189–199. https://doi.org/10.1111/ldrp.12141

Shalev, R. S., Auerbach, J., Manor, O., & Gross-Tsur, V. (2000). Development dyscalculia: Prevalence and prognosis. *European Child and Adolescent Psychiatry, 9*, II/58–II/64. https://doi.org/10.1007/s007870070009

Shalev, R. S., Manor, O., & Gross-Tsur, V. (2005). Developmental dyscalculia: A prospective six-year follow-up. *Developmental Medicine and Child Neurology, 47*, 121–125. https://doi.org/10.1017/s0012162205000216

Stock, P., Desoete, A., & Roeyers, H. (2010). Detecting children with arithmetic disabilities from kindergarten: Evidence from a 3-year longitudinal study on the role of preparatory arithmetic abilities. *Journal of Learning Disabilities, 43*(3), 250–268. https://doi.org/10.1177/0022219409345011

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

Szűcs, D., & Goswami, U. (2013). Developmental dyscalculia: Fresh perspectives. *Trends in Neuroscience and Education, 2*, 33–37. https://doi.org/10.1016/j.tine.2013.06.004

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 de Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2019). *Elementary and middle school mathematics: Teaching Developmentally* (10th edition). Pearson.

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

van Lieshout, E. C. D. M., & Xenidou-Dervou, I. (2018). Pictorial representations of simple arithmetic problems are not always helpful: A cognitive load perspective. *Educational Studies in Mathematics, 98*, 39–55. https://doi.org/10.1007/s10649-017-9802-3

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM Mathematics Education, 52*, 1–16. https://doi.org/10.1007/s11858-020-01130-4

Wang, A. Y., Fuchs, L. S., & Fuchs, D. (2016). Cognitive and linguistic predictors of mathematical word problems with and without irrelevant information. *Learning and Individual Differences, 52*, 79–87. https://doi.org/10.1016/j.lindif.2016.10.015

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

What Works Clearinghouse (2020). *Procedures handbook, version 4.1.* Institute of Education Sciences.

Wei, X., Lenz, K. B., & Blackorby, J. (2013). Math growth trajectories of students with disabilities: Disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education*, *34*(3), 154–165. https://doi.org/10.1177/0741932512448253

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://doi.org/10.1037/0022-0663.80.2.192

Xin, Y. P. (2008). The effect of schema-based instruction in solving mathematics word problems: An emphasis on prealgebraic conceptualization of multiplicative relations. *Journal for Research in Mathematics Education*, *39*(5), 526–551.

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

	PMEQ Individual[a]				PMEQ Small-Group			
	PMEQ (*n* = 105)		BaU (*n* = 115)		PMEQ (*n* = 56)		BaU (*n* = 20)	
	n	%	*n*	%	*n*	%	*n*	%
Gender (female)	62	59.0	67	58.3	30	53.6	12	60.0
Race/ethnicity								
African American	13	12.4	12	10.4	8	14.5	5	25.0
Hispanic/Latino/Latina	79	75.0	82	71.3	35	61.8	9	45.0
Caucasian	4	3.6	6	5.2	8	14.5	3	15.0
Asian	4	3.6	3	2.6	1	1.8	2	10.0
Multi-racial	5	4.8	7	6.1	1	1.8	1	5.0
Other	3	2.9	4	3.5	3	5.5	0	0.0
Students in special education	18	16.2	11	9.6	9	16.3	3	15.0
Dual-language learners	64	61.0	68	59.1	30	53.6	7	35.0

Note. BaU = Business as usual; PMEQ = Pirate Math Equation Quest.

[a]Data from Powell et al. (2021)

Table 2

Daily Session Content for Individual PMEQ and Small-Group PMEQ Interventions

Individual PMEQ Intervention		Small-Group PMEQ Intervention	
Session	Content	Session	Content
1-4	Addition/subtraction computation review	1-3	Addition/subtraction computation review
5	Introduce Total schema	4	Introduce Total schema
6-7	Total schema	5-10	Total schema
8-10	Addition/subtraction computation review		
11-15	Total schema		
16	Addition/subtraction computation review		
17	Introduce Difference schema	11	Introduce Difference schema
18-33	Total and Difference schemas	12-18	Total and Difference schemas
34	Introduce Change schema	19	Introduce Change schema
35-42	Total, Difference, and Change schemas	20-27	Total, Difference, and Change schemas

		28	Introduce Equal Groups schema
43-51	Review of Total, Difference, and Change schemas	29-39	Review of Total, Difference, Change, and Equal Groups schemas

Table 3

Correlations Between Measures

	1	2	3	4	5	6
1. Screening Single-Digit Word Problems	--					
2. Screening Texas Word Problems-Brief	.248	--				
3. Pretest Texas Word Problems-Part 1 and Part 2	.353	.483	--			
4. Pretest STAAR-Part 1 and Part 2	.249	.446	.793	--		
5. Posttest Texas Word Problems-Part 1 and Part 2	.352	.424	.673	.605	--	
6. Posttest STAAR-Part 1 and Part 2	.257	.350	.612	.607	.850	--

Note. All correlations significant at $p < .05$.

Table 4

PMEQ Small Groups: Student-Level and Classroom-Level Descriptive Statistics

	Pretest			Posttest		
	n	*M*	*SD*	*n*	*M*	*SD*
Student-level						
Proximal word problem						
BaU	20	3.55	2.46	20	7.50	5.77
PMEQ	56	7.14	5.21	48	17.48	8.11
Distal word problem						
BaU	20	0.85	0.88	20	1.85	1.84
PMEQ	56	2.14	2.12	48	5.46	4.18
Classroom-level						
Proximal word problem						
BaU	5	3.55	1.97	5	7.50	5.39
PMEQ	14	7.14	3.68	14	17.95	6.56
Distal word problem						
BaU	5	0.85	0.45	5	1.85	1.10
PMEQ	14	2.14	1.39	14	5.60	3.67

Note. BaU = Business as usual; PMEQ = Pirate Math Equation Quest.

Table 5

PMEQ Small Groups: Multilevel Analysis Estimating the Main Effect of Intervention

	Proximal word problem				Distal word problem			
	Estimate	*SE*	*p*-value	ES [95% CI]	Estimate	*SE*	*p*-value	ES [95% CI]
Intercept	10.71	2.15	0.00		3.34	1.33	0.03	
Pretest Level 1	0.75	0.18	0.00		0.90	0.17	0.00	
Pretest Level 2	1.21	0.32	0.00		1.34	0.49	0.01	
PMEQ	5.67	2.58	0.04	0.76 [.05, 1.46]	1.85	1.36	0.19	0.51 [-.24, 1.25]
	Variance	ICC			Variance	ICC		
Level 1 (student)	19.06	0.56			3.69	0.37		
Level 2 (classroom)	14.67	0.44			4.12	0.42		
Level 3 (school)					2.10	0.21		

Note. PMEQ = Pirate Math Equation Quest.

Table 6

PMEQ Small Groups Versus PMEQ Individual: ANCOVA Results Estimating the Main Effect of Intervention

	PMEQ Small Group				PMEQ Individual			
	Estimate	*SE*	*p*-value	ES [95% CI]	Estimate	*SE*	*p*-value	ES [95% CI]
Intercept	10.08	1.40	0.00		6.73	0.51	0.00	
Pretest	0.97	0.16	0.00		0.59	0.11	0.00	
PMEQ	6.18	1.71	0.00	0.83 [0.57, 1.10]	11.45	0.76	0.00	1.99 [1.82, 2.16]

Note. PMEQ = Pirate Math Equation Quest.