Schema-Based Word-Problem Intervention with and without Embedded Language Comprehension Instruction

Lynn S. Fuchs[1], Pamela M. Seethaler[1], Sonya K. Sterba[1], Caitlin Craddock[1], Douglas Fuchs[1], Donald L. Compton[2], David C. Geary[3], and Paul Changas[4]

[1]Vanderbilt University, [2]Florida State University, [3]University of Missouri, [4]Metropolitan-Nashville Public Schools

Inquiries should be sent to Lynn S. Fuchs, Vanderbilt University, 228 Peabody, 110 Magnolia Circle, Suite 418, Nashville, TN 37203.

This research was supported by 2 R01 HD053714 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to Vanderbilt University in the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health.

Abstract

The main purposes of this randomized controlled trial were to (a) investigate the efficacy of schema-based word-problem (WP) intervention at first grade, (b) assess the added value of embedding language comprehension (LC) instruction within schema-based WP intervention, and (c) examine whether number knowledge intervention, with its effects on arithmetic but without explicit WP intervention, are sufficient to address the WPS deficits of at-risk first graders. We randomly assigned 1st-grade children (n=391; mean age = 6.53 years *SD* = 0.32) to 4 conditions: WP schema-based intervention with embedded LC instruction, the same schema-based WP intervention without LC instruction, number knowledge intervention, and control. The 3 intervention conditions included 45 30-min sessions. Multilevel models, accounting for classroom and school effects, revealed added value for embedded LC instruction to improve WPS, and WP-language partially mediated this effect between the 2 WP intervention conditions. Both schema-based WP intervention conditions outperformed number knowledge intervention and control, and number knowledge intervention with its effects on arithmetic skill was insufficient to outperform the control group on WP.

KEY WORDS: math word problems, word-problem development, word-problem intervention, language comprehension

Schema-Based Word-Problem Intervention

with and without Embedded Language Comprehension Instruction

Word-problem solving (WPS) is a strong school-age predictor of employment and wages in adulthood (Every Child a Chance Trust 2009; Batty, Kivimäki, & Deary, 2010). It reflects understanding of and the capacity to apply mathematical ideas in everyday life and in the service of science, technology, engineering, and advanced mathematics learning. WP difficulty is also widespread (Doroczy, Wolska, Meurers, & Nuerk, 2015) and can occur with intact arithmetic skill (Cummins, Kintsch, Ruesser, & Wiener, 1988; Fuchs et al., 2008; Koedinger & Nathan, 2004). Specific WP difficulty may occur because WPS engages more cognitive resources than does arithmetic (Doroczy et al.; Swanson & Beebe-Frankenberger, 2004). A notable distinction between the two forms of mathematical cognition is that WP relies more on language comprehension (LC) than does arithmetic (Fuchs et al., 2016, 2018; Cummins, 1991; Singer, Strausser, & Cuadro, 2019).

Schema-based intervention is a validated method for improving WP performance among struggling learners. With schema-based intervention, children learn to classify word problems into problem types and then solve problems according to structured solution methods for each problem type. In multiple randomized controlled trials, schema-based intervention has proven efficacious for enhancing WP performance (e.g., Fuchs et al., 2008; Fuchs, Fuchs, Finelli, Courey, & Hamlett, 2004; Jitendra, Star, Rodriguez, Lindell, & Someki, 2011; Jitendra et al., 2009), but its value has previously not been assessed at first grade, where the dominant intervention focus for at-risk learners is number knowledge and arithmetic.

The main purposes of this randomized controlled trial were to (a) investigate the efficacy of schema-based WP intervention at first grade, (b) assess the added value of embedding LC instruction within schema-based WP intervention, and (c) examine whether number knowledge intervention, with its effects on arithmetic but without explicit WP intervention, are sufficient to address the WPS deficits of at-risk first graders. We defined *risk* as low arithmetic skill and low performance (below the 25th percentile) on math concepts and applications. Also note that LC standard scores for the study sample fell 1.67 standard deviation (*SD*) below the test's normative framework.

Students were randomly assigned, stratifying by classroom, to four conditions: schema-based WP intervention with embedded LC instruction, the same schema-based WP intervention without LC instruction, number knowledge intervention, and control (the schools' business-as-usual program, including classroom instruction and, for most students, school-based intervention). Each of the three active intervention conditions lasted 15 weeks (three sessions per week, 30 min per session) and included five min of speeded, strategic arithmetic practice as conventionally done within schema-based WP intervention (see for examples the Fuchs et al. line of studies).

Method

Participants

We conducted this study in accordance with our university-approved IRB protocol. We screened 3,009 consented children in 186 classrooms in 21 schools to enroll 417 at-risk and 480 not-at-risk children in this study. Teachers administered two alternate-form practice screening tests (*First-Grade Test of Mathematics Computation* and *First-Grade Test of Concepts and Applications*; Fuchs, Hamlett, & Fuchs, 1990) within 1 week of the study's screening session, which was conducted by research assistants. Established cut-scores were applied to identify at-risk children (combined score across screeners < 15, with scores < 4 on Computation and < 12 on Concepts and Applications). This corresponds to a percentile cut-score of < 25th percentile. To avoid false positives, we excluded 14 students whose teachers identified them as entirely or almost entirely non-English speakers. Because the study's interventions were not designed to address the needs of students with intellectual disability, we excluded 23 students with standard scores below 80 on both subtests of the 2-subtest Wechsler Abbreviated Intelligence Scale (WASI; Wechsler, 2011).

We randomly sampled 417 of the remaining children, ensuring no more than five at-risk participants in the same classroom. We then randomly assigned students, at the individual level while stratifying by classroom, to the four study conditions: number knowledge intervention (NK), WP intervention (WP), WP intervention with embedded WP-language instruction (WP[L]), and control. Over the course of participation, 25 at-risk children (10 NK, 6 WP, 3 WP[L], six control) moved beyond the study's reach, leaving a final sample of 391 at-risk children.

Screening and other descriptive data are provided by at-risk conditions in Table 1. At-risk children in the NK, WP, WP[L], and control conditions, respectively, were 42%, 37%, 40%, and 39% male. They were 43%, 40%, 33%, and 35% African American, 16%, 21%, 23%, and 20% white non-Hispanic, 33%, 32%, 38%, and 38% white Hispanic, and 8%, 7%, 6%, and 7% other. They were 34%, 34%, 41%, and 38% English-learners, with 79%, 78%, 80%, and 75% from economically disadvantaged households. There were no significant differences among the four at-risk conditions. Not-at-risk classmates were 43% male; 30% African American, 40% white non-Hispanic, 20% white Hispanic, and 10% other. Thirty-four percent were English-learners; 19% received subsidized lunch. Thus, compared to not-at-risk classmates, at-risk children (across conditions) were disproportionately African American and white Hispanic and from economically disadvantaged households. There were no significant difference among at-risk conditions on any screening and other descriptive measure.

Also note that a **sample of not-at-risk classmates was also included to serve as a comparison group for judging the severity of at-risk students' pretest performance gaps and posttest achievement gaps**. **Not-at-risk classmates were <u>not</u> involved in the randomized**

control trial. To identify not-at-risk classmates, we randomly sampled 480 children from among those (a) scoring *above* the cut-point on each math screener and with a combined score > 17, (b) scoring *above* the established risk cut-point (19) on *Word Identification Fluency* (Fuchs, Fuchs, & Compton, 2004) to avoid the complication of not-at-risk students with reading difficulty, (c) who were entirely or almost entirely non-English speakers, and (d) scoring *at or above* 80 on both WASI subtests. Over the course of participation, 25 not-at-risk children moved to schools outside the study's reach, leaving a final sample of 455 not-at-risk classmates. Not-at-risk children were assessed on the same measures completed by at-risk participants. See Table 1 for not-at-risk screening and other descriptive data. Not-at-risk classmates outperformed at-risk students on each measure. Performance gaps between at-risk children to not-at-risk classmates (effect sizes [ESs]; Hedges *g*, 1982) are provided in Table 2.

Screening Measures

First-Grade Test of Computational Fluency (Fuchs et al., 1990) is a 25-item test that samples the typical first-grade computation curriculum: adding two single-digit numbers (9 items), subtracting two single-digit numbers (10 items), adding three single-digit numbers (2 items), adding two 2-digit numbers without regrouping (2 items), and subtracting a 1-digit number from a 2-digit number (2 items). Students have 2 min to complete as many items as possible. Staff entered responses into a computerized program on an item-by-item basis, with 15% of tests re-entered by an independent scorer. Data-entry agreement was 99.6. Sample-based α was .98.

First-Grade Test of Mathematics Concepts and Applications (Fuchs et al., 1990) is a 25-item test sampling the typical first-grade concepts/applications curriculum (i.e., numeration, concepts, geometry, measurement, applied computation, money, charts/graphs, WPs). The tester reads the words in each item aloud. For 20 items, students have 15 s to respond; for 5 items, 30 s.

Staff entered responses into a computerized program on an item-by-item basis, with 15% re-entered by an independent scorer. Data-entry agreement was 98.5%. Sample-based α was .93.

Word Identification Fluency (Fuchs et al., 2004) provides students with 1 min to read a list of 50 words randomly sampled from 100 high-frequency pre-primer, primer, and first-grade words. If a student finishes before 1 min, the score is prorated. We administered two alternate forms and averaged scores. Alternate-form reliability/stability is .97; correlations with Woodcock Reading Mastery Test-Word Identification (Woodcock, 1998) are .77-.82.

WASI (Wechsler, 2011) is a 2-subtest measure of general cognitive ability, comprising *Vocabulary* and *Matrix Reasoning* subtests (reliability > .92). *Vocabulary* assesses expressive vocabulary, verbal knowledge, memory, learning ability, and crystallized and general intelligence. Students identify pictures and define words. *Matrix Reasoning* measures nonverbal fluid reasoning and general intelligence. Students complete matrices with missing pieces.

Other Descriptive Measures

Language. *Woodcock Diagnostic Reading Battery* (*WDRB) - Listening Comprehension* (Woodcock, 1997) measures the ability to understand sentences or passages. With 38 items, students supply the word missing at the end of sentences or passages that progress from simple verbal analogies and associations to discerning implications. Reliability is .80.

Nonverbal reasoning. *WASI Matrix Reasoning* (Wechsler, 2011) measures nonverbal reasoning with pattern completion, classification, analogy, and serial reasoning tasks. Students complete a matrix, from which one section is missing, from five response options. Reliability is .94.

Working memory. The *Working Memory Test Battery for Children – Listening Recall* and *Counting Recall* (WMTB-C; Pickering & Gathercole, 2001), both dual-task subtests, each has six items at span levels from 1-6 to 1-9. Passing four items at a level moves the child to the next level. At each span level, the number of items to be remembered increases by one. Failing three items terminates the subtest. Subtest order is designed to avoid overtaxing any component area and is generally arranged from easiest to hardest. We used the trials correct score. For *Listening Recall*, children determine if a sentence is true; after making true/false determinations for a series of sentences, they recall the last word of each sentence. For *Counting Recall*, children count a set of 4, 5, 6, or 7 dots on a card; after counting a series of cards, they recall the number of counted dots on each card.

Attentive behavior. The Strength and Weaknesses of ADHD-Symptoms and Normal-Behavior (*SWAN*; Swanson et al., 2004) samples items from the *Diagnostic and Statistical Manual of Mental Disorders-IV* criteria for Attention Deficit Hyperactivity Disorder for inattention (9 items) and hyperactivity-impulsivity (9 items), but scores are normally distributed. Teachers rate items on a 1-7 scale. We report data for the inattentive subscale as the average rating across the nine items. The SWAN correlates well with other dimensional assessments of behavior related to attention (www.adhd.net). Sample-based α was .99.

Mathematics Measures

Arithmetic. From the *First-Grade Mathematics Assessment Battery* (Fuchs, Hamlett, & Powell, 2003), *Arithmetic Combinations* includes two subtests. Addition comprises 25 addition problems with sums from 5 to 12 (two items have an addend of 1; one has an addend of zero). Subtraction comprises 25 subtraction fact problems with minuends from 5 to 12 (one item has a minuend of 1; one has a minuend of zero). For each subtest, students have 1 min to write answers. Because the pattern of results was similar across subtests, we used total number of correct answers across addition and subtraction. Sample-based α was .96.

Word problems. At pretest, we administered *Pennies Story Problems* (Jordan & Hanich, 2000), which comprises 14 WPs representing combine, compare, or change schema. The scenarios all involve pennies. Problems require addition and subtraction (sums and minuends to 12) for solution. The tester reads each item aloud; students have 30 s to write an answer and can ask for re-reading(s) as needed. The score is the number of correct number answers (*pennies* is the correct label for every problem). Sample-based α was .86.

At posttest, we administered *First-Grade Word Problems* (Fuchs et al., 2009), which includes 12 WPs representing combine, compare, and change schemas, with/without irrelevant information, relevant quantities presented in graphs, superordinate terms in questions, tricky questions (no reference to what the problem is mostly about or to associated problem-type vocabulary), implicit change verbs, non-compare *–er* words, and unusual time passages conjunctions. One combine problem involves three parts. Solutions require addition and subtraction arithmetic (sums and minuends to 12). Testers read a WP aloud; students follow along on paper, with up to 2 min to write the answer before testers read the next problem. Each problem is scored for correct math (1 point) and label (1 point) to reflect processing of the WP statement and understanding of the problem's theme. Sample-based α was .86.

Word-problem language. *Word Problem-Language Assessment* (Fuchs, Craddock, & Seethaler, 2013) includes two subtests. On each, testers read WPs aloud while the child follows along on paper. The child can request re-readings as needed. For each problem, the child decides whether addition or subtraction is needed to solve the problem and selects from three choices to indicate which words are most important in determining whether to add or subtract. Testing begins with a practice item, with which testers explain what they want the child to do. Testers point to the problem as they read: "Lacy has 1 pink flower and 2 yellow flowers. How many flowers does she have altogether? Here's my first question: To find the answer to this problem, do you add or subtract (tester points to the words, *add* and *subtract*)? Let's think. If we want to know how many flowers she has altogether, we add the pink flowers plus the yellow flowers to find the answer. So you say, 'add.' Here's my second questions, 'Which words are the most important to help you figure out if you add or subtract: *altogether* (tester points) or *How many* (tester points) or *1 pink flower* (tester points)? The word *altogether* is most important in telling us to add. So you say, 'altogether.'" For each problem, testers pose the questions used for the practice questions, point to answers to help the child focus on the response options, and record answers on a score sheet. There are 12 problems representing combine, compare, and change schemas, with/without irrelevant information, superordinate terms in questions, implicit change verbs, non-compare *–er* words, and unusual time passages conjunctions. Each WP earns two possible points, for a maximum score of 24. Sample-based α was .74.

Intervention

When describing interventions, we use the present tense because this study's interventions still exist. When describing study conditions, we use the past tense to communicate those procedures are completed.

Commonalities across conditions. The three intervention conditions shared four commonalities. First, intervention comprised 45 30-min sessions conducted one-to-one over 15 weeks outside the classroom in the child's school. Absences and snow days were made up to ensure each child received 45 sessions.

Second, instruction was explicit, designed to compensate for the domain-general cognitive and linguistic limitations associated with WP difficulty (Kintsch & Greeno, 1985; Fuchs et al., submitted). Research syntheses indicate the importance of explicit instruction for improving at-risk children's learning (Baker et al., 2002; Gersten et al., 2009). In the three interventions, explicit instruction (a) ensures students have the foundational knowledge and skills to succeed with new content; (b) provides explanations in simple, direct language; (c) models efficient solution strategies instead of expecting students to discover strategies on their own;

gradually fades support for correct execution of taught strategies; (e) provides interleaved practice so students use knowledge and strategies to generate many correct responses and distinguish among problem types; and (h) incorporates systematic cumulative review.

Third, because at-risk first graders often display attention, motivation, and self-regulation difficulties that may affect learning (e.g., Fuchs et al., 2013), intervention includes a self-regulation system centered on four rules: use inside voice; stay in seat; follow directions; and try hard to answer problems correctly. Tutors set a timer to beep at 5-min intervals. Whenever the timer beeps, tutors award a checkmark if the child is following all four rules. Tutors keep track of checkmarks earned. At the end of the session, checkmarks are converted to stickers on the child's chart. When the sticker chart is full (~weekly), the child picks a small prize.

The third commonality across conditions was that each session comprises three segments: speeded practice on arithmetic problems (5 min); the lesson, which tutors introduce and systematically review concepts and strategies (20 min); and practice (5 min). Fourth, throughout these segments, tutors require children to know the answer or use taught counting strategies to solve arithmetic problems.

With *speeded practice*, which is referred to as Meet or Beat Your Score, children have 60s to answer flash cards. In the first six lessons, cards are restricted to n+/-1, n+/-0, and n+/-2; after efficient addition and subtraction counting strategies are taught, all combinations of addends and minuends up to 18 are included. Children are taught to "know the answer right off the bat" (retrieve from memory) if confident; otherwise, use the taught counting strategies. Children answer each presented problem correctly because, as soon as an error occurred, the tutor requires them to use the taught counting strategy to produce the correct response. To discourage guessing or careless application of counting strategies, seconds elapse as children execute the counting strategy as many times as needed to produce the correct answer. In this way, careful but quick responding increases the number of correct responses. Children have a chance to meet or beat the first score, and the day's higher score is plotted.

To teach *counting strategies*, tutors first address the conceptual bases using manipulatives and the number line and then teach how to use fingers to execute these strategies. For addition, children "count-in." For 3 + 4 =, they hold 3 fingers up on one hand to represent the smaller quantity; then they put 4 remaining fingers down in a fist to note the 4 in hand and count 5 (putting another finger down with the other down fingers), 6 (putting another finger down), and 7 (putting the last finger down). The last number counted is the answer. For subtraction, they "count-up." For 5 – 2 = 3, they count the difference between the numbers, saying 2 and then they count 3 (hold up a finger), 4 (hold up a finger), 5 (hold up a finger). The number of raised fingers is the answer.

Number knowledge intervention. To foster engagement, number knowledge intervention, known as *Galaxy Math* (previously *Number Rockets*), incorporates a space theme (e.g., rocket ship Meet or Beat Your Score charts; moon beam manipulatives; Fuchs et al., 2013; see Fuchs, Fuchs, Craddock, & Seethaler, 2019 for a complete manual). After speeded practice, tutors conduct the lesson, relying on number lines and manipulatives to represent mathematical ideas (1-19 for Units 1-5; 1-100 for Unit 6). See Table S1 for an overview of lesson content, organized into units.

Unit 1 (lessons 1-4) addresses basic number knowledge. Unit 2 (lessons 5-6) focuses on adding and subtracting concepts and principles (e.g., adding and subtracting as moving up and down the number line; +0, +1, and +2 problems as simple counting on; the meaning of the equal sign; commutative property of addition; inversion principle). Unit 3 (lessons 7-11) teaches counting strategies. Unit 4 (lessons 12-13) focuses on doubles concepts. Unit 5 (lessons 14-37) focuses on number sets 5 - 12 (e.g., the 5 set includes all problems with sums or minuends of 5).

Unit 6 (lessons 38-45) focuses on writing, counting and reading numbers 0-99, 3-addend problems, and double-digit adding and subtracting.

Units 4 and 5, which focus on partitioning numbers into constituent sets (for the 5 set: 0 + 5, 1 + 4, 2 + 3, 5 – 0, 5 - 1, 5 - 2, etc.) and number families (four number sentences with the same three numbers), comprise most of the program. Three lessons are allocated to each set. Five activities are conducted in each lesson. (1) Children use unifix cubes to explore how the target number in that set is partitioned in different ways to derive the adding and subtracting problems comprising the set. (2) Children use blocks and visual displays that group families in the set to strengthen part-whole understanding. (3) Children generate all addition and subtraction problems with answers in the target set, using manipulatives to represent problems. (4) Children answer problems in previous sets, with corrective feedback. (5) Children and tutors play a game with a space theme using the number set for that day. For example, a game board with cartoon aliens and footprints stretches to end of a game board. For each problem answered correctly, children move one footprint forward on the board. Tutors and children compete against each other, with tutors occasionally providing incorrect answers and children executing correction procedures.

The 5-min *practice* segment of each lesson is a Bingo game. Children complete arithmetic problems, covering answers with rocket ships on a card displaying nine numbers. Winning Bingo earns them an extra checkmark. The winning Bingo rule varies (e.g., 3 in a row, 4 corners, cover all). Tutors require counting strategies on errors to produce correct answers.

Word-problem solving intervention, with and without embedded LC instruction. To foster engagement, WP intervention, known as *Pirate Math*, incorporates a pirate theme (pirate-themed sticker charts; gold coin manipulatives, Find x!; see Fuchs, Fuchs, & Seethaler, 2015 for a complete manual). After speeded practice, tutors conduct the lesson. For an overview of lesson content, see Tables S2 and S3, respectively, for WP intervention with and without embedded language instruction.

Content is organized in five units. Unit 1 (lessons 1-9) addresses adding and subtracting concepts, addition and subtraction counting strategies, and solving for a missing number, represented by the letter x. Note that there is substantially less focus on number knowledge, including adding/subtracting and arithmetic concepts and principles, in the WP than the number knowledge condition. Unit 2 (lessons 10-18) focuses on total problems (combining two or three quantities to make a total; e.g., There are 5 girls on the playground and 3 girls in the yard. How many girls are there?). Unit 2 also includes instruction on 3-addend addition and 3-part total problems. Unit 3 (lessons 19-27) focuses on difference problems (comparing a larger and a smaller quantity to find the difference; e.g., At the picnic, the kids ate 5 hotdogs. They ate 3 hamburgers. How many more hot dogs did they eat than hamburgers?). Unit 4 (lessons 28-36) focuses on change problems (increasing or decreasing a start quantity to produce an end quantity; e.g., Jamarius baked 6 cookies. Then, he gave 3 of them to his friend. How many cookies does Jamarius still have?). Unit 5 (lessons 37-45) introduced a sorting game where students decide whether a problem is total, difference, or change; it also provides review and practice.

Units 2-4 begin by teaching the mathematical structure of the focal WP type for that unit. This involves role playing the problem type's central mathematical event using an intact number story (no missing quantity), concrete objects, and the child's and tutor's names. Tutors next use the intact story to connect the mathematical central event to (a) a visual schematic (into which story quantities are written) and (b) a hand gesture, which is used across lessons to quickly remind children of the three problem types' central events. Then tutors connect the problem's central event to a problem-model number sentence: for combine (referred to as *total*): P1 + P2 = T (Part 1 plus Part 2 equals Total; also 3-part problems); for compare (referred to as *difference*): B – s = D (bigger quantity minus smaller quantity equals difference); for change: St +/- C = E (for change increase, start number plus change number equals end number; for change decrease, start number minus change number equals end number).

Tutors finally introduce a problem (with a missing quantity), using the same cover story with which the lesson is introduced. The problem is enacted via role playing with concrete objects and the child's/tutor's names; the problem type's schematic and hand gesture are applied; and the problem model number sentence is introduced with x standing for the missing quantity. In each WP unit, the idea of irrelevant information is taught.

In Units 2-4 for both WP conditions, tutors explicitly teach step-by-step strategies to reduce demands on reasoning and working memory. This includes strategies for understanding WPs as belonging to WP types (schemas or problem models) and for building the WP model. As children process and solve a WP, they use these strategies to name the problem type, to represent that problem model with the model number sentence, to enter relevant quantities from the WP statement into the problem model while crossing out "extra" (irrelevant) numbers, and to solve for the missing quantity.

Children RUN through the problem: Read it, Underline the segment in which the problem's object code (which becomes the label) is revealed, and Name the problem type. They write T, D, or C next to the problem to help them remember the problem type, and they write the problem type's model number sentence. Then, they re-read the problem as they enter known quantities and x to stand for the unknown quantity into the slots of the problem type's number sentence. For example, given the total problem, There are 5 girls on the playground and 3 girls in the yard. There are also 4 boys in the classroom. How many girls are there?, the tutor reads the problem as the child follows along. The child underlines *girls*; identifies the problem as total and writes T and the problem type's model number sentence, P1 + P2 = T; re-reads the problem while replacing 5 for P1, 3 for P2, and x for T; and crosses out 4. To solve 5 + 3 = ___, the child retrieves the answer or uses counting strategies. Go to S4 for photos of a first grader solving a total problem.

The embedded language comprehension component addresses WP language relevant to the first-grade WP schemas. Some are specific to problem types. In Unit 2, the meaning and application of vocabulary are taught in the context of combine problems: joining words (e.g., *altogether, in all*) and superordinate categories (e.g., *animals* = *dogs* + *cats*). In Unit 3, the meaning and application of vocabulary and syntax are taught in the context of compare problems: compare words (e.g., *more, fewer, than, -er* words) and adjective *-er* versus verb *-er* words (e.g., *bigger* vs. *teacher*). In change problems, the meaning and application of vocabulary and syntax are taught in the context change problems: cause - effect conjunctions (e.g., *then, because, so*), implicit quantity change verbs (e.g., *cost, ate, found*), and time passage phrases (e.g., *3 hours later, the next day*). Other focal points apply across problem types: for example, confusing cross-problem constructions (e.g., *more than* vs. *then … more*) and "tricky" labels (e.g., questions with superordinate category words, without a label, noun that's the wrong label [as in money questions]).

It is important to note that embedded language instruction does *not* teach children to rely on key words. Instead, children are taught that processing words in isolation is an error-fraught strategy; are taught to read problems so they can meaningfully derive the problem's central mathematical event; and are taught specifically how and why "grabbing numbers and key words" to identify a number sentence frequently produces wrong answers. To help children appreciate this, they check the work of "other children" (prepared worked problems). Students find errors and explain how and why errors occurred. Worked examples rely on key words to select the wrong operation or misuse irrelevant numbers or fail to recognize 3-part total problems.

The 5-min *practice* segment of each lesson provides students independent practice finding the unknown (x) in number sentences and solving a WP. WPs align with that day's lesson focus or provide a review problem. Tutors provide corrective feedback as needed.

Training and support for tutors. Across cohorts, 54 full or part-time tutors employed by the project served as tutors. Almost all were university master's students (7% in teacher licensure programs; 5% licensed before graduate school). Each tutor worked with 5-6 students, distributed equally across the two intervention conditions.

Tutors were introduced to the intervention program in a 2-day workshop and then supported to implement the program in weekly meetings throughout the 15 weeks of intervention. The initial workshop provided an overview; explained the program's focus and rationale; discussed distinctions between the intervention conditions and methods to ensure each student received the condition consistent with his/her random assignment condition; modeled key elements of the intervention; provided practice in implementing these elements; and explained methods for providing appropriate corrective feedback throughout lessons. After the workshop but prior to the first lesson with a student, tutors completed a reliability quiz covering major components of the intervention program with 90% accuracy and then demonstrated at least 90% accurate implementation of lesson components with one of two project coordinators. To promote fidelity, tutors also studied lesson scripts to guide their teaching, but were not permitted to read or memorize scripts.

During implementation, tutors attended weekly meetings, in which tutors provided updates on their students, discussed learning and behavior challenges, and problem solved with each other, with the two project coordinators, and with the first author. Key information on upcoming topics were also reviewed and upcoming materials were distributed. Also, every intervention session was audiotaped. Staff listened to a randomly selected lesson for each tutor in each condition on a weekly basis. The purpose was to identify difficulties with or deviations from the protocol and provide corrective feedback as needed.

Fidelity of Implementation

These audiotapes were also used to quantify fidelity of implementation. Research assistants independently listened to tapes while completing a checklist to identify the percentage of essential points addressed in that lesson. We sampled 15% of tapes such that intervention conditions, tutors, and lesson types were sampled comparably. Coding agreement exceeded 96%. The mean percentage of points addressed was 97.70 (*SD*=2.96) in NK; 97.66 (*SD*=2.33) in WP; and 96.80 (*SD*=2.98) in WP[L]. Thus, fidelity was high and similar across conditions.

Procedure

In August, we screened students for study entry. In September - October, we administered descriptive measures individually and mathematics measures in small groups. In November, teachers completed the SWAN. Intervention began in late October and continued through March. After intervention ended, we administered mathematics measures in small groups and the WP language test individually. All test sessions were audiotaped; 15% of tapes were randomly selected, stratifying by tester, for accuracy checks by an independent scorer. Agreement exceeded 99%. Testers were blind to study condition when administering and scoring tests.

Data Analysis

The study's data structure involved three levels of nesting: 391 students nested within 186 classrooms nested within 21 schools. Three-level multilevel models (e.g., Raudenbush & Bryk, 2002) were employed to account for nesting for the arithmetic and WP-language outcomes, but only two-level models were estimable for the WPS outcome. All models were fit in M*plus* 7.4 (Muthen & Muthen, 1998-2019) using full-information robust maximum likelihood estimation (MLR).

Our first set of research questions involved the main effect of intervention condition on arithmetic outcome, controlling for pretest arithmetic; the main effect of intervention condition on WPS outcome controlling for pretest story problems; and the main effect of intervention condition on WP language. These questions were evaluated using the multilevel models in

Equations (1) - (3).

$$\begin{aligned} poWP_{ij} &= \beta_{0j} + \beta_{1j}\mathrm{D1}_{ij} + \beta_{2j}\mathrm{D2}_{ij} + \beta_{3j}\mathrm{D3}_{ij} + \beta_{4j}\, prWP_{ij} + e_{ij} \\ \beta_{0j} &= \gamma_{00} + u_{0j}; \beta_{1j} = \gamma_{01}; \beta_{2j} = \gamma_{02}; \beta_{3j} = \gamma_{03}; \beta_{4j} = \gamma_{04} \end{aligned} \tag{1}$$

$$\begin{aligned} poWP[L]_{ijk} &= \beta_{0jk} + \beta_{1jk}\mathrm{D1}_{ijk} + \beta_{2jk}\mathrm{D2}_{ijk} + \beta_{3jk}\mathrm{D3}_{ijk} + e_{ijk} \\ \beta_{0jk} &= \gamma_{000} + u_{0jk} + u_{00k} \\ \beta_{1jk} &= \gamma_{001}; \beta_{2jk} = \gamma_{002}; \beta_{3jk} = \gamma_{003} \end{aligned} \tag{2}$$

$$\begin{aligned} poARI_{ijk} &= \beta_{0jk} + \beta_{1jk}\mathrm{D1}_{ijk} + \beta_{2jk}\mathrm{D2}_{ijk} + \beta_{3jk}\mathrm{D3}_{ijk} + \beta_{4jk}\, prARI_{ijk} + e_{ijk} \\ \beta_{0jk} &= \gamma_{000} + u_{0jk} + u_{00k} \\ \beta_{1jk} &= \gamma_{001}; \beta_{2jk} = \gamma_{002}; \beta_{3jk} = \gamma_{003}; \beta_{4jk} = \gamma_{004} \end{aligned} \tag{3}$$

where: $e_{ijk} \sim N(0, \sigma^2);\ u_{0jk} \sim N(0, \tau_{00}^{(2)});$ and $u_{00k} \sim N(0, \tau_{00}^{(3)})$

Here, *i* indexes student, *j* indexes classroom, and *k* indexes school. e_{ijk} is a student-level residual with variance σ^2, and u_{0jk} and u_{00k} are, respectively, classroom-level and school-level deviations from the average intercept γ_{000} with variances of $\tau_{00}^{(2)}$ and $\tau_{00}^{(3)}$, respectively. $prSD_{ij}$ is pretest story problems and $prBF_{ijk}$ is pretest arithmetic. VPS is WPS outcome; poOL is WP language; poBF is posttest arithmetic. $\mathrm{D1}_{ijk}, \mathrm{D2}_{ijk}, \mathrm{D3}_{ijk}$ are dummy variables representing, respectively, NK intervention versus control, WP intervention versus control, and WP[L] intervention versus control. We tested the difference between WP intervention and WP[L] by testing $(\gamma_{003} - \gamma_{001}) = 0$; the difference between WP intervention and arithmetic intervention by testing $(\gamma_{002} - \gamma_{001}) = 0$; and the difference between the two WP conditions by testing $(\gamma_{002} - \gamma_{001}) = 0$.

An additional research question involved the indirect effect of intervention contrasts on the outcomes via WP language, controlling for pretest story problems and pretest arithmetic. This was investigated using the multilevel mediation model in Equation (4).

$$\begin{aligned}
&poWP[L]_{ijk} = \beta^{p}_{0jk} + \beta^{p}_{1jk}\mathrm{C1}_{ijk} + e^{p}_{ijk} \\
&\quad \beta^{p}_{0jk} = \gamma^{p}_{000} + u^{p}_{0jk} + u^{p}_{00k}; \beta^{p}_{1jk} = \gamma^{p}_{001}; \\
&poWP_{ijk} = \beta^{v}_{0jk} + \beta^{v}_{1jk}\mathrm{C1}_{ijk} + \beta^{v}_{2jk}\, poWP[L]_{ijk} + \beta^{v}_{3jk}\, prWP_{ijk} + \beta^{v}_{4j}\, prARI_{ijk} + e^{v}_{ijk} \\
&\quad \beta^{v}_{0jk} = \gamma^{v}_{000} + u^{v}_{0jk}; \beta^{v}_{1jk} = \gamma^{v}_{001}; \beta^{v}_{2jk} = \gamma^{v}_{002}; \beta^{v}_{3jk} = \gamma^{v}_{003}; \beta^{v}_{4jk} = \gamma^{v}_{004}
\end{aligned} \tag{4}$$

$$e^{p}_{ijk} \sim N(0, \sigma^{2p});\; e^{v}_{ijk} \sim N(0, \sigma^{2v})$$

$$\text{where: } \begin{bmatrix} u^{p}_{0jk} \\ u^{p}_{00k} \\ u^{v}_{0jk} \end{bmatrix} \sim MVN\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau^{p(2)}_{00} & 0 & 0 \\ 0 & \tau^{p(3)}_{00} & 0 \\ 0 & 0 & \tau^{v(2)}_{00} \end{bmatrix} \right)$$

Parameters and residuals in the equation for the mediator (WP language) are denoted with a superscript "*p*." Parameters and residuals in the equation for the WPS outcome are denoted with a superscript "*v*." $\mathrm{C1}_{ijk}$ is a dummy code for a specific comparison only: NK intervention versus control or WP intervention versus control or WP[L] intervention versus control or WP intervention versus WP[L] intervention. Hence, four versions of Equation (4) were fit with different definitions of $\mathrm{C1}_{ijk}$ to evaluate four different indirect effects. Defining the *a*-path as $\hat{\gamma}^{p}_{001}$ and the *b*-path as $\hat{\gamma}^{v}_{002}$, the point estimate for the indirect effect in each case was $\gamma^{p}_{001} \times \gamma^{v}_{002}$ (i.e., *a*-path × *b*-path), and its confidence interval (CI) was obtained using Preacher and Selig's (2012) Monte Carlo method, as implemented at www.quantpsy.org.

Results

For the three outcomes (arithmetic, WPS, WP language), intraclass correlations (ICCs) at the classroom level were .022, .010, and .036, respectively. ICCs at the school level were .062, .001, and .020, respectively. See Table 1 for means and *SD*s by at-risk conditions and for not-at-risk classmates. See Table 2 for at-risk children's pre- and post-intervention performance gaps (against not-at-risk classmates) on the outcome measures (Hedges *g* ESs). See Table 3 for results of main effects multilevel models in Equations (1) - (3), as well as Benjamini-Hochberg procedure critical values and Hedges g ESs between at-risk conditions.

On the arithmetic outcome, effects of all three active conditions, controlling for pretest arithmetic, were not significantly different from each other but were significantly stronger than control. For the main effect of intervention on WPS, controlling for pretest story problems, the effect of WP[L] intervention was significantly stronger than WP intervention, which was significantly stronger than arithmetic intervention, which was not significantly different from control. On WP-language, the effect of WP[L] intervention was significantly stronger than that of WP intervention, which was significantly stronger than arithmetic intervention, although WP intervention and arithmetic intervention were not significantly different from control. As shown in Table 1, significant effects were retained after the Benjamini-Hochberg procedure (Thissen, Steinberg, & Kuang, 2002) was applied to control for Type 1 error.

For the multilevel mediation models, the indirect effect for the contrast between NK intervention versus control on WPS outcome, via WP-language, was not significant, $\hat{\gamma}^{p}_{001} \times \hat{\gamma}^{v}_{002}$ = -0.163 CI = {-.571, .095} (*a*-path $\hat{\gamma}^{p}_{001}$ = -.505(.443) *p*= .254; *b*-path $\hat{\gamma}^{v}_{002}$ = .323(.095) *p* = .001; direct effect or c'-path $\hat{\gamma}^{v}_{001}$ = .259(.319) *p* = .417, and total effect or *c*-path = .060(.363) *p* = .869). The indirect effect for the contrast between WP intervention versus control on WPS outcome, via WP-language, also was not significant, $\hat{\gamma}^{p}_{001} \times \hat{\gamma}^{v}_{002}$ =.296 CI = {-.042, .679} (*a*-path $\hat{\gamma}^{p}_{001}$ = .925(.529) *p* = .081; *b*-path $\hat{\gamma}^{v}_{002}$ = .320(.110) *p* = .004; direct effect or c'-path $\hat{\gamma}^{v}_{001}$ = 4.637(.476) *p* < .001; and total effect or *c*-path = 4.763(.490) *p* < .001).

By contrast, both the indirect effects involving WP[L] intervention on WPS outcome via WP language were significant: for the contrast between WP[L] intervention versus control, $\hat{\gamma}^{p}_{001} \times \hat{\gamma}^{v}_{002}$ =.668 CI = {.326, 1.058} (*a*-path $\hat{\gamma}^{p}_{001}$ = 2.412(.544) *p* < .001; *b*-path $\hat{\gamma}^{v}_{002}$ = .277(.065) *p* < .001; direct effect or c'-path $\hat{\gamma}^{v}_{001}$ = 6.791(.584) *p* < .001, and total effect or *c*-path = 7.246(.581) *p* < .001), and for the contrast between WP[L] intervention versus WP intervention, $\hat{\gamma}^{p}_{001} \times \hat{\gamma}^{v}_{002}$ = .554

CI = {.138, 1.147} (*a*-path $\hat{\gamma}^{p}_{001}$ = 1.601(.477) $p < .001$; *b*-path $\hat{\gamma}^{v}_{002}$ = .346(.113) $p = .002$; direct effect or c'-path $\hat{\gamma}^{v}_{001}$ = 1.872(.637) $p = .003$; and total effect or *c*-path = 2.470(.653) $p < .001$).

Discussion

Does Embedding Language Comprehension Instruction within Schema-Based Intervention Provide Added Value over Schema-Based intervention without Language Comprehension Intervention?

The first conclusion is that embedding LC instruction within WP intervention (WP[L]) produces stronger WPS development than does the same WP intervention without LC instruction. The ES for this contrast was almost one-half standard deviation (*SD*). This is considerable given that intervention time was held constant across the two WP conditions, thus resulting in less direct skills WP instruction than was provided in WP intervention without LC instruction (approximately 5 min less per session; 225 min less across 45 sessions). So although both WP intervention conditions significantly enhanced WPS over the control group (ES = 1.75 for WP[L] intervention; ES = 1.08 for WP intervention), the WP[L] condition produced considerably stronger WP performance than did WP intervention without a LC component. This demonstrates the added value of LC instruction and indicates LC's causal contribution to WPS.

This finding is clarified by the pattern of effects involving WP language. In the main effects model on the WP language outcome, only the WP[L] condition outperformed the control group. ESs were 0.17 for NK versus control, 0.16 for WP versus control, and 0.56 for WP[L]. More pertinent to the present study's purpose, the mean difference between the two WP conditions was significant with an ES of 0.41. Moreover, the mediation analysis, testing whether the effects between the two WP conditions on the WPS outcome are derived at least partly via children's WP language learning, revealed a significant indirect effect.

It is instructive to consider the paths that contribute to this indirect effect. The significant *b*-path reflects the previously established relation between WP-LC and WPS (Fuchs et al., 2018), in which children with stronger WP language demonstrate stronger WPS. The significant *a*-path reflects WP[L] intervention's added value over the same WP intervention without LC instruction on WP language. The product of these *a*- and *b*-paths produces a significant indirect effect. This indicates that WP[L] intervention improves WPS *directly*, as revealed in the significant c' (direct) effect, via an indirect process by which WP[L] *strengthens WP language, which further improves WPS*. This mediation effect was also specific: The indirect effect via WP language was not significant for the NK versus control group contrast or the WP versus control group contrast. It *was* significant for the WP[L] versus control group contrast. The significant indirect effects involving WP[L] were partial, such that the direct and total effects of WP[L] intervention over WP intervention were also significant.

In terms of embedded LC instruction's added value, it is interesting to note at-risk children's post-intervention WP achievement gaps. Over the course of intervention, the gap substantially narrowed in the WP intervention without LC instruction condition: from 1.69 *SD*s below not-at-risk classmates at pre-intervention to 0.11 at post-intervention. Even so, in the WP[L] condition, the gap more than closed, decreasing from 1.65 *SD*s below not-at-risk classmates before intervention to 0.52 *SD*s *above* not-at-risk classmates afterwards. This reflects WP[L] students' added boost of competence, via improved WP language, hopefully providing an additional layer of protective cushion as they exit supplemental math intervention. This underscores the value of embedding WP-LC instruction within WP intervention.

Finding that LC plays a causal role in WPS has implications for designing WP intervention and for supporting children's language development. With respect to language development, results suggest that parents and preschool teachers be vigilant for and act on opportunities to explicitly extend children's ordinary word usage to mathematical contexts. In terms of WP instruction, results indicate that WP intervention should incorporate a deliberate focus on language, including but not limited to WP-specific vocabulary and syntactic knowledge (e.g., understanding the distinction between *more than* and *then there were more*; that the cause and effect in change WPs may be presented in either order within WP statements).

Is Transfer from Improved Arithmetic Skill Insufficient to Address WP Learning Challenges?

Our second conclusion addresses the question, Is transfer from enhanced arithmetic skill, achieved via number knowledge intervention, sufficient to address the WP learning challenges of at-risk learners? The present randomized control trial indicates the answer is no. Although NK intervention did substantially improve arithmetic skill, with an ES of 0.59 over at-risk control group students, this did not translate into stronger WP performance. The ES for NK intervention on WPS was 0.09, and the pre- to posttest WP gap for NK intervention remained sizeable: 1.70 SDs below not-at-risk classmates at pretest and 1.34 at posttest. This is a similar to the control group's pre- and post WP achievement gap (1.76 to 1.27 *SD*s.)

Finding an absence of transfer from arithmetic skill to WPS echoes Fuchs et al. (2014), in which classwide calculations instruction combined with small-group intervention for at-risk second graders improved calculations performance but failed to transfer to WPS. Results of that study focused on outcomes across learner types. Powell, Powell, Fuchs, Cirino, Fuchs, Compton, and Changas (2015), who focused specifically on at-risk learners, documented the same phenomenon. Those analyses, combined with the present study's corroborating results, substantiate the need for explicit WP intervention.

Understanding the necessity of a deliberate focus on WPS for at-risk learners is important because teachers often view WPs as arithmetic tasks (Doroczy et al., 2015). Findings should alert teachers, colleges of education, and textbook developers to the need for research-based WP instructional methods. Unfortunately, WP instruction in the U.S. is commonly fraught with errorproducing methods by teaching children to link keywords to mathematical operations and by providing children practice that requires application of the same solution method across all WPs on the practice sheet. These methods discourage the mathematical reasoning required for WPs (Powell & Fuchs, 2018).

Finding an absence of transfer also extends understanding about and provides theoretical insight into the nature of mathematics competence: how different aspects of mathematical cognition relate to each other, which aspects of performance are shared or distinct, and how difficulty in one domain corresponds to difficulty in another. But few studies have assessed transfer from calculations to WPS. The present study adds to that small body of literature by providing corroborating evidence at a different grade level that arithmetic and WPs represent different components of mathematical performance. Results also explain why WP difficulty may occur with intact arithmetic skill.

Does Schema-Based WP Instruction Improve WP Outcomes at First Grade

Our third conclusion concerns the efficacy of schema-based WP instruction. Prior work has not been conducted at first grade. Superior performance for both WP conditions over the control group was expected given prior research on schema-based WP instruction at other grade levels (Fuchs et al., 2008; Fuchs et al., 2004; Jitendra et al., 2011; Jitendra et al., 2009). This expectation was supported with both versions of WP intervention, each of which relied on schema-based instruction. We therefore conclude that schema-based WP intervention is efficacious for at-risk children as young as 5-7 years of age.

References

Aiken, L. R. (1972). Language factors in learning mathematics. *Review of Educational Research, 42,* 359-385. doi:10.3102/00346543042003359

Anderson, U. (2007). The contribution of WM to children's mathematics word problem solving. *Applied Cognitive Psychology, 21*, 1201-1216.

Baker, S., Gersten, R., & Lee, D-S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal*, 103, 51-73.

Batty, G. D., Kivimäki, M., & Deary, I. J. (2010). Intelligence, education, and mortality. *BMJ: British Medical Journal*, *340*, 1-2. doi: http://dx.doi.org/10.1136/bmj.c563

Bernardo, A. B. I. (1999). Overcoming obstacles to understanding and solving word problems in mathematics. *Educational Psychology, 19*, 149–163.

Bull, R., & Johnston, R. S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology, 65*, 1-24.

Catts, H. W., Hogan, T. P., & Adolf, S. M. (2005). Developmental changes in reading and reading disabilities. In H. W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 25-40). Mahwah, NJ: Erlbaum.

Catts, H. W., & Kamhi, A. G. (2017). Prologue: Reading comprehension is not a single ability. *Language, Speech, and Hearing Services in Schools, 48*(2), 73-76.

Cummins, D. D. (1991). Children's interpretations of arithmetic word problems. *Cognition and Instruction, 8*, 261-289, DOI: 10.1207/s1532690xci0803_2

Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, *20*, 405-438. https://doi.org/10.1016/0010-0285(88)90011-4

Daroczy, G., Wolska, M., Neurers, W. D., & Nuerk, H.-C. (2015). Word problems: A review of linguistic and numerical factors contribution to their difficulty. *Frontiers in Psychology, 6*, 348. doi: 10.3389/fpsyg.2015.00348.

Davis-Dorsey, J., Ross, S. M., & Morrison, G. R. (1991). The role of rewording and context personalization in the solving of mathematical word problems. *Journal of Educational Psychology, 83*, 61-68.

De Corte, E., & Verschaffel, L. (1988). Computer simulation as a tool in research on problem solving in subject-matter domains. *International Journal of Educational Research, 12,* 49-69.

Dehaene, S., Spelke, E. S., Pinel, P., Stanescu-Cosson, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284*(5416), 970-974. doi:10.1126/science.284.5416.970

Every Child a Chance Trust (2009). *The long-term costs of numeracy difficulties*. Retrieved. August 14, 2009, http://www.everychildachancetrust.org/counts/index.cfm.

Fitzmaurice, G. M., Laird, N. M., &Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed.). New York: Wiley.

Fuchs, L.S., Craddock, C., & Seethaler, P.M. (2013). *Word Problem-Language Assessment*. Available from L.S. Fuchs, Vanderbilt University, Nashville, TN 37203.

Fuchs, L.S., Fuchs, D., Craddock, C., & Seethaler, P.M. (2019). *Galaxy Math Teachers Manual*. Available from L.S. Fuchs, Vanderbilt University, Nashville, TN 37203.

Fuchs, L.S., Fuchs, D., & Compton, D.L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*, 7-21.

Fuchs, L.S., Fuchs, D., Compton, D.L., Hamlett, C.L., & Wang, A.Y. (2015). Is word-problem solving a form of text comprehension? *Scientific Studies of Reading*, *19*, 204–223. http://dx.doi.org/10.1080/10888438.1005745

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C., & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98,* 29-43.

Fuchs, L. S., Fuchs, D., Finelli, R., Courey, S. J., & Hamlett, C. L. (2004). Expanding schema-based transfer instruction to help third graders solve real-life mathematical problems. *American Educational Research Journal, 41*, 419-445.

Fuchs, L. S., Fuchs, D., Seethaler, P. M., & Barnes, M. A. (submitted). The role of working memory in mathematical word-problem solving: Implications for instruction and intervention. *ZDM Mathematics Education*.

Fuchs, L.S., Fuchs, D., Stuebing, K., Fletcher, J.M., Hamlett, C.L., & Lambert, W.E. (2008). Problem-solving and computation skills: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology, 100,* 30-47.

Fuchs, L.S., Geary, D.C., Compton, D.L., Fuchs, D., Hamlett, C.L., Seethaler, P.M., Bryant, J.V., & Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical and general cognitive abilities? *Developmental Psychology, 46,* 1731-1746. doi: 10.1037/a0020662

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*, 58-77.

Fuchs, L.S., Gilbert, J.K., Fuchs, D., Seethaler, P.M., & Martin, B. (2018). Text comprehension and oral language as predictors of word-problem solving: Insights into word-problem solving as a form of text comprehension. *Scientific Studies of Reading, 22,* 152-166. https://doi.org/10.1080/10888438.2017.1398259

Fuchs, L.S., Gilbert, J.K., Powell, S.R., Cirino, P.T., Fuchs, D., Hamlett, C.L., Seethaler, P.M., & Tolar, T.M. (2016). The role of cognitive processes, foundational math skill, and calculation accuracy and fluency in word-problem solving versus pre-algebraic knowledge. *Developmental Psychology, 52*, 2085-2098. doi: 10.1037/dev0000227

Fuchs, L.S., Hamlett, C.L., & Fuchs, D. (1990). *First Grade Test of Computational Fluency and First-Grade Test of Mathematics Concepts and Applications*. Available from L.S. Fuchs, 228 Peabody, Vanderbilt University, Nashville, TN 37203.

Fuchs, L.S., Hamlett, C.L., & Powell, S.R. (2003). *Second-Grade Calculations* Battery. Available from L. S. Fuchs, Vanderbilt University, Nashville, TN 37203.

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*, 990-1006. DOI: 10.1037/a0036793.

Fuchs, L.S., Seethaler, P.M., Craddock, C., & Fuchs, D. (2019). Pirate Math Grade 1 Manual. Available from L.S. Fuchs, Vanderbilt University, Nashville, TN 37203.

Fuchs, L.S., Seethaler, P.M., & Craddock, C. (2009). First-Grade Math Word Problems. Available from L.S. Fuchs, Vanderbilt University, Nashville, TN 37203.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*, 1202–1242. doi: 10.3102/0034654309334431

Gough, P.B., & Tunmer, W.E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.

Haskell, R. E. (2001). *Transfer of learning: Cognition, instruction, and reasoning*. Academic.

Hecht, S.A., Torgesen, J.K., Wagner, R.K., & Rashotte, C.A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology, 79*, 192-227.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92*, 490-499. http://dx.doi.org/10.1037/0033-2909.92.2.490

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* (retrieved 3/26/2019). https://www.researchgate.net/profile/Eric_Hedberg/publication/228513219_Final_report_on_the_national_survey_of_algebra_teachers_for_the_National_Math_Panel/links/55364e710cf20ea35f11ca6d/Final-report-on-the-national-survey-of-algebra-teachers-for-the-National-Math-Panel.pdf.

Jitendra, A. K., Star, J. R., Rodriguez, M., Lindell, M., & Someki, F. (2011). Improving students' proportional thinking using schema-based instruction. *Learning and Instruction, 21*, 731-745.

Jitendra, A. K., Star, J. R., Starosta, K., Leh, J. M., Sood, S., Caskie, G., Hughes, C. L., & Mack, T. R. (2009). Improving seventh grade students' learning of ratio and proportion: The role of schema-based instruction. *Contemporary Educational Psychology, 34*, 250-264.

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

Ke, Z., & Wang, L. (2015). Detecting individual differences in change: Methods and comparisons. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*, 382-400.

Kintsch, W., & Greeno, J. G. (l985). Understanding and solving word arithmetic problems. *Psychological Review, 92*, 109-129.

Lee, K., Ng, S. F., Ng, E. L., & Lim, Z. Y. (2004). Working memory and literacy as predictors of performance on algebraic word problems. *Journal of Experimental Child Psychology, 89*, 140-158.

Meneghetti, C., Carretti, B., & De Beni, R. (2006). Components of reading comprehension and scholastic achievement. *Learning and Individual Differences, 16*, 291-301.

Melby-Lervag, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*, 270-291. doi:10.1037/a0028228.

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nathan, M.J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognitive and Instruction, 9*, 392-389.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school*. The National Academies Press.

Peng, P., Namkung, J., Barnes, M., & Sun, C. (in press). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*. Advance online publication. http://dx.doi.org/10.1037/edu0000079

Pickering, S., & Gathercole, S. (2001). *Working Memory Test Battery for Children.* London: The Psychological Corporation.

Powell, S.R., & Fuchs, L.S. (2018). Effective word-problem instruction: Using schemas to facilitate mathematical reasoning. *TEACHING Exceptional Children, 51*(1), 31-42. PMID 30210179; PMC 6130842; doi.org/10.1177/0040059918777250

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

Powell, S. R., Driver, M. K., Roberts, G., & Fall, A.-M. (2017). An analysis of the mathematics vocabulary knowledge of third- and fifth-grade students: Connections to general vocabulary and mathematics computation. *Learning and Individual Differences, 57*, 22–32. doi:10.1016/j.lindif.2017.05.011

Purpura, D. J., & Ganley, C. M. (2014). Working memory and language: Skill-specific or domain-general relations to mathematics? *Journal of Experimental Child Psychology 122*, 104–121.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, 20, 110-122.

Riley, M.S., Greeno, J. G., & Heller, J. H. (1983). Development of children's problem-solving ability in arithmetic. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp.153 – 196). San Diego, CA: Academic.

Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly*, *23*(2), 139-159, DOI: 10.1080/10573560601158461

Schumacher, R. F., & Fuchs, L. S. (2012). Does understanding relational terminology mediate effects of intervention on compare word problems? *Journal of Experimental Child Psychology, 111*, 607-628.

Singer, V., Strausser, K., & Cuadro, A. (2019). Direct and indirect paths from arithmetic to school performance. *Journal of Educational Psychology, 111*, 434-445. http://dx.doi.org/10.10737/edu0000290

Stern, E., & Lehrndorfer, A. (1992). The role of situational context in solving word-problems. *Cognitive Development, 7*, 259-268.

Swanson, J. et al. (2004). Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and the SWAN rating scales. Downloaded from www.adhd.net on 12/20/2004.

Swanson, H. L. (2016). Word problem solving, working memory and serious math difficulties: Do cognitive strategies really make a difference? *Journal of Applied Research in Memory and Cognition, 5*, 368-383. DOI: 10.1016/j.jarmac.2016.04.012

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 96*, 471-491. DOI: 10.1037/0022-0663.96.3.471.

Swanson, H. L., Cooney, J. B., & Brock, S. (1993). The influence of working memory and classification ability on children′s word problem solution. *Journal of Experimental Child Psychology, 55*, 374-395.

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

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

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology*, *79*, 294-321.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedures for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics, 27*, 77-83.

Ukrainetz, T. A. (2017). Commentary on "Reading Comprehension Is Not a Single Ability:" Implications for child language intervention. *Language, Speech, and Hearing Services in Schools, 48*(2), 92-97.

Van der Schoot, M., Bakker Arkema, B. A. H., Horsley, T. M., & Van Lieshout, E. D. C. M. (2009). The consistency effect depends on markedness in less successful but not successful problem solvers: An eye movement study in primary school children. *Contemporary Educational Psychology, 34*, 58–66. http://dx.doi.org/10.1016/ j.cedpsych.2008.07.002

van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: Academic.

Vicente, S., Orarntia, J., & Verschaffel, L. (2007). Influence of situational and conceptual rewording on word problem solving. *British Journal of Educational Psychology, 77,* 829-848. DOI:10.1348/000709907X178200

Vilenius-Tuohimaa, P. M., Aunola, K., & Nurmi, J. E. (2008). The association between mathematical word problems and reading comprehension. *Educational Psychology, 28,* 409-426.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.

Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., Defries, J. C., Olson, R. K., & Pennington, B. F. (2013). Comorbidity between reading disability and math disability: Concurrent psychopathology, functional impairment, and neuropsychological functioning. *Journal of Learning Disabilities*, *46*, 500-516. doi: 10.1177/0022219413477476.

Woodcock, R. W. (1997). *Woodcock Diagnostic Reading Battery*. Itasca, IL: Riverside.

Table 1
Means and Standard Deviations by Study Condition

	At-Risk Study Condition									
	Control (n=104)		Number Know (n=92)		Word Problem (n=96)		Word Problem [L] (n=99)		Not-At-Risk (n=455)	
Variable	Mean	*(SD)*	Mean	*(SD)*	Mean	*(SD)*	Mean	*(SD)*	Mean	*(SD)*
Screening										
Concepts, Application, Calculations	10.00	(6.13)	9.63	(6.66)	9.59	(5.96)	8.72	(5.09)	27.38	(10.05)
Word Identification Fluency	14.05	(17.13)	15.56	(17.22)	12.71	(15.65)	13.70	(15.34)	43.74	(23.10)
WASI: Vocabulary	37.89	(9.51)	36.88	(9.84)	37.08	(9.76)	36.68	(9.95)	47.33	(11.42)
Matrix Reasoning	45.26	(6.53)	43.15	(4.79)	44.95	(5.45)	44.03	(7.02)	50.44	(7.71)
Descriptive										
Listening Comprehension	77.70	(16.39)	75.57	(18.63)	75.14	(17.79)	73.64	(19.42)	94.01	(16.88)
Working Memory: Listen Recall	67.21	(15.50)	70.13	(17.39)	66.48	(15.50)	67.36	(16.21)	92.20	(19.64)
Count Recall	80.55	(15.40)	78.15	(13.47)	80.16	(14.79)	78.04	(13.73)	93.33	(15.73)
Key Math-Problem Solving	93.37	(5.89)	92.72	(5.26)	93.07	(5.87)	92.78	(5.54)	104.96	(7.93)
WRAT-Arithmetic	88.41	(10.96)	88.07	(10.15)	89.38	(11.11)	88.11	(11.88)	110.76	(11.04)
Outcomes										
Arithmetic: Pre	5.85	(4.77)	6.45	(5.20)	5.90	(4.13)	5.61	(3.91)	18.24	(9.27)
Post	16.16	(11.47)	24.32	(13.33)	24.01	(12.47)	24.87	(11.41)	35.69	(14.94)
Word Problems: Pre	1.55	(1.20)	1.33	(1.23)	1.54	(1.26)	1.62	(1.28)	6.27	(3.02)
Post	3.36	(2.38)	3.47	(2.98)	7.40	(5.12)	9.81	(4.94)	7.82	(3.50)
Word-Problem Language: Post	13.41	(3.79)	13.07	(4.02)	13.07	(3.69)	15.55	(3.82)	18.24	(3.35)

Table 2

Performance Gaps by Study Condition: Effect Sizes (ESs; Hedges g) for At-Risk Study Conditions versus Not-At-Risk Classmates

	At-Risk Study Condition			
	Control (n=104)	Number Know (n=92)	Word Problem (n=96)	Word Problem [L] (n=99)
Variable	ES	ES	ES	ES
Screening				
Concepts, Application, Calculations	1.84	1.86	1.88	1.99
Word Identification Fluency	1.34	1.27	1.41	1.37
WASI: Vocabulary	0.85	0.93	0.92	0.95
Matrix Reasoning	0.69	1.00	0.74	0.84
Descriptive				
Listening Comprehension	0.97	1.07	1.11	1.17
Working Memory: Listen Recall	1.32	1.15	1.36	1.30
Count Recall	0.82	0.99	0.85	0.99
Key Math-Problem Solving	1.52	1.61	1.56	1.61
WRAT-Arithmetic	2.03	2.08	1.93	2.02
Outcomes				
Arithmetic: Pre	1.44	1.35	1.43	1.48
Post	1.36	0.77	0.80	0.75
Word Problems: Pre	1.70	1.76	1.69	1.65
Post	1.34	1.27	0.11	-0.52
Word-Problem Language: Post	1.41	1.49	1.51	0.78

Table 3

Main Effects Multi-Levels Model Results (n=391)

Model/Parameter	Estimate	SE	*p*-value	B-H critical	Hedges *g* ES
Arithmetic					
Fixed Effects					
Intercept	8.675	1.049	<.0001		
Adjusted Mean Difference					
NK v. control	7.419	1.528	<.0001	.0125	0.59
WP v. control	7.985	1.475	<.0001	.0083	0.65
WP[L] v. control	9.154	1.272	<.0001	.0042	0.79
WP v. NK	0.566	1.882	.764	.0167	0.03
WP[L] v. NK	1.735	1.496	.246	.0208	0.14
WP v. WP[L]	1.169	1.345	.385	.0250	0.10
Pretest effect	1.332	0.110	.001		
Variance Components					
Student-level residual	99.374	6.385			
Classroom-level intercept	3.367	7.133			
School-level intercept	6.426	6.071			
Word Problems					
Fixed Effects					
Intercept	2.383	0.360	<.0001		
Adjusted Mean Difference					
NK v. control	0.260	0.401	.517	.0250	0.09
WP v. control	4.045	0.526	<.0001	.0125	1.08
WP[L] v. control	6.416	0.550	<.0001	.0042	1.75
WP v. NK	3.785	0.576	<.0001	.0167	0.94
WP[L] v. NK	6.156	0.563	<.0001	.0083	1.55
WP v. WP[L]	2.371	0.684	.0001	.0208	0.47
Pretest effect	0.632	0.171	<.0001		
Variance Components					
Student-level residual	14.746	1.280			
Classroom-level intercept	0.715	1.395			
Word-Problem Language					
Fixed Effects					
Intercept	13.511	0.355	<.0001		
Mean Difference					
NK v. control	-0.386	0.413	.349	.0250	0.17
WP v. control	0.573	0.504	.255	.0208	0.16
WP[L] v. control	2.164	0.559	<.0001	.0125	0.56
WP v. NK	0.960	0.386	.013	.0167	0.24
WP[L] v. NK	2.550	0.331	<.0001	.0042	0.63
WP v. WP[L]	1.590	0.444	<.0001	.0083	0.41
Variance Components					
Student-level residual	13.064	1.084			
Classroom-level intercept	1.180	1.034			
School-level intercept	0.283	0.315			

Note. *p*-values for *z*-tests of variance components are conservative and thus often not reported (divide est/SE and compare to +1-1.96 to discern significance according to z-test). An adjustment Fitzmaurice et al. (2011, p. 209) suggest is to employ α = .10 rather than α = .05; it performs similarly to other correction methods (Ke & Want, 2015). B-H is the Benjamini-Hochberg procedure critical value. ES is effect size.

Figure 1.

Models assessing the indirect effect of WP language on the WPS outcome for four intervention contrasts: NK vs. control (Panel A), WP v. control (Panel B), WP[LC] vs. control (Panel C), and most central to the present study's purpose, WP[LC] vs. WP (Panel D).