Published in final edited form as:
J Educ Psychol. 2014 November ; 106(4): 990–1006.

Does Calculation or Word-Problem Instruction Provide A Stronger Route to Pre-Algebraic Knowledge?

Lynn S. Fuchs[1], **Sarah R. Powell**[2], **Paul T. Cirino**[3], **Robin F. Schumacher**[1], **Sarah Marrin**[1], **Carol L. Hamlett**[1], **Douglas Fuchs**[1], **Donald L. Compton**[1], and **Paul C. Changas**[4]

[1]Vanderbilt University

[2]University of Texas at Austin

[3]University of Houston

[4]Metropolitan Nashville Public Schools

Abstract

The focus of this study was connections among 3 aspects of mathematical cognition at 2nd grade: calculations, word problems, and pre-algebraic knowledge. We extended the literature, which is dominated by correlational work, by examining whether intervention conducted on calculations or word problems contributes to improved performance in the other domain and whether intervention in either or both domains contributes to pre-algebraic knowledge. Participants were 1102 children in 127 2nd-grade classrooms in 25 schools. Teachers were randomly assigned to 3 conditions: calculation intervention, word-problem intervention, and business-as-usual control. Intervention, which lasted 17 weeks, was designed to provide research-based linkages between arithmetic calculations or arithmetic word problems (depending on condition) to pre-algebraic knowledge. Multilevel modeling suggested calculation intervention improved calculation but not word-problem outcomes; word-problem intervention enhanced word-problem but not calculation outcomes; and word-problem intervention provided a stronger route than calculation intervention to pre-algebraic knowledge.

Mathematics, which involves the study of quantities as expressed in numbers or symbols, comprises a variety of related branches. At the primary grades, the major curricular focus is whole numbers, which is conceptualized in three domains: understanding number, calculations, and word problems. In the intermediate grades and middle school, the next major curricular topics are rational numbers and algebraic thinking, each of which includes its own domains. In high school, curriculum offerings include algebra, geometry, trigonometry, and calculus. Little is understood, however, about how such aspects of mathematical cognition relate to each other: which aspects of performance are shared or

Inquiries should be sent to: Lynn S. Fuchs, 228 Peabody, Vanderbilt University, Nashville, TN 37220; lynn.fuchs@vanderbilt.edu.

[1]Our use of terms is as follows. *Mathematical equations* are those with one variable (whereas *algebraic equations* have two or more variables). *Standard equations* are those with operations on the left side followed by the equal sign and either a single value (the total sum) or an unknown quantity (x or blank) at the end of the problem. Thus, 3 + __ = 7 is not a nonstandard equation because the operation is on the left and the sum is on the right. However, 7 = 3 + __ is nonstandard, as is __ = 3 + 4.

distinct; how difficulty in one domain corresponds to difficulty in another; or whether instruction in one or another domain produces better learning in a third domain. Such understanding would provide theoretical insight into the nature of mathematics competence and practical guidance about how to organize curriculum and design instruction. The focus of the present study was connections among three aspects of mathematics performance in second-grade children: arithmetic calculations, arithmetic word problems, and pre-algebraic knowledge.

Focus on Arithmetic Calculations, Arithmetic Word Problems, and Pre-Algebraic Knowledge

Few studies have examined how students develop competence with algebra. Yet, consensus exists that algebra is required for successful participation in the workforce and represents a gateway to higher forms of learning in mathematics, science, technology, and engineering (National Mathematics Advisory Panel [NMAP], 2008; RAND Mathematics Study Panel, 2003). For these reasons, passing an algebra course is frequently required for high-school graduation, but 35% of students fail to complete such a course and 93% of 17-year-olds cannot solve multistep algebra problems (U.S. Department of Education, 2008). In light of such difficulty, interest in algebraic cognition among elementary grade children has increased in the past decade, with the 2003 RAND report calling for systematic inquiry on this topic – hence our focus on pre-algebraic knowledge.

At the same time, calculations and word problems are, in and of themselves, critical aspects of mathematics competence in the primary grades and through adulthood. Whereas a calculation problem is set up for solution, a word problem requires students to process text to build a problem model and construct a number sentence for calculating the unknown. This transparent difference would seem to alter the nature of the task, and correlational studies suggest the cognitive abilities underlying word problems and calculations differ (e.g., Fuchs, Fuchs, Stuebing, et al., 2008; Fuchs, Geary, et al., 2010a, b; Swanson, 2006).

Although such correlational work raises the possibility that calculations and word problems represent distinct domains of mathematical cognition, stronger evidence would come from studies examining whether intervention in one domain affects the other. A handful of experimental studies suggest limited transfer from calculation intervention to word-problem outcomes (e.g., Fuchs et al., 2009; Fuchs, Powell, et al., 2011). But we identified no studies assessing transfer from word-problem intervention to calculation outcomes, none investigating both forms of transfer in the same study design, and none exploring transfer from calculations or word-problem intervention to pre-algebraic knowledge. In the present study, we extended the literature by examining whether intervention conducted on calculations or word problems transfers to the other domain and whether intervention in either or both domains contributes to pre-algebraic knowledge.

The Role of Arithmetic Calculations and Word Problems in Algebraic Thinking

Algebra involves symbolizing and operating on numerical relationships and mathematical structures. Algebraic expressions can be treated procedurally, by substituting numerical values to yield numerical results (Kieran, 1990). This suggests that understanding of arithmetic principles involves generalizations that are algebraic in nature, such that algebra warrants a prominent role in early instruction (Blanton & Kaput, 2005; Carraher & Schliemann, 2002; Falkner, Levi, & Carpenter, 1999; NMAP, 2008). Others (e.g., Balacheff, 2001; Linchevski, 2001) suggest an interference effect that makes algebra developmentally inappropriate for young children: Sfard (1991) referred to a *deep ontological gap* (p. 4); Linchevski and Herscovics (1996) used the term *cognitive gap* (p. 39). Many researchers, however, espouse a third view, closely connected to the first, which supports a connection, but only if arithmetic instruction (on calculations or word problems) is designed to facilitate the transition (Herscovics & Kieran, 1980; Jacobs, Franke, Carpenter, Levi, & Battey, 2007).

Underpinning the third perspective is Pillay, Wilss, and Boulton-Lewis's (1998) model of learning, which incorporated the work of Kieran (1990) to reinterpret the cognitive *cut* or *gap* in terms of a developmental progression. In this model, the first stage of learning is arithmetic competence: the capacity to operate numerically and the understanding of operational laws and relational meaning of the equal sign (i.e., both sides of the equal sign are the same value) in standard equations. This provides the foundation for a pre-algebraic stage, which builds on arithmetic competence by expanding the relational meaning of the equal sign to include nonstandard equations (i.e., equations with the equal sign in an unfamiliar, or nonstandard, position), the concept of unknowns in equations, and the concept of a variable. This stage supports the development of formal algebraic competence. Pillay et al.'s model is consistent with a connection between algebra and arithmetic, as expressed in the first perspective discussed above. But it is more central to the third perspective by specifying an intermediary stage that clarifies the nature of that connection, even as the model captures focal points for measuring the progression from arithmetic to algebra and designing early instruction to support the transition.

Unfortunately, research illustrates how conventional instruction causes most students to misconstrue the equal sign as an operational symbol (Baroody & Ginsburg, 1983; McNeil & Alibali, 2005; Powell, 2012). Take the problem $7 + 5 = \_\_ + 3$. Common errors reflect misunderstanding about the equal sign: 12, in which students ignore the operation to the right side of the equal sign, and 15, in which they add all known values (Falkner, Levi, & Carpenter, 1999). Such confusion persists into high school (NMAP, 2008) and is associated with difficulty in using algebraic notation to represent word problems (Powell & Fuchs, 2010) and solve linear equations (Alibali, Knuth, Hattikudar, McNeil, & Stephens, 2007; Knuth, Stephens, McNeil, & Alibali, 2006). In the present study, we relied on Pillay et al.'s model to design both forms of intervention (calculations and word problems) in ways that potentially support the transition from arithmetic to algebra. (We provide the theoretical and empirical basis for the intervention design in the method section.)

Fluency with calculations may reflect a strong foundation in arithmetic operational laws and generalizations, which may support pre-algebraic thinking. For example, Jacobs et al. (2007) showed positive effects of a year-long professional development project, in which algebraic reasoning as generalized arithmetic was used as a platform for work with elementary teachers. Students in participating classes showed stronger understanding of the equal sign and used more strategies involving relational thinking during interviews than students in nonparticipating classes. Fluency with calculations may also support pre-algebra by reducing demands on working memory to free up attention for the challenges associated with handling nonstandard equations and variables (Geary et al., 2008).

On the other hand, word problems not only require calculations, but also involve two forms of symbolic representations (numerals and language) and reflect understanding of relationships between known and unknown quantities. In fact, a key source of error in word problems involves transforming problem narratives into algebraic equations (Geary et al., 2008). For example, competent problem solvers translate “Fred had 3 more than Harry” to F = 3 + H, by recognizing the smaller quantity must increase to equal the larger quantity. A common error, however, is F + 3 = H. Word problems may, therefore, involve greater symbolic complexity than calculations and may rely more on the type of mental flexibility, manipulation of symbolic associations, and maintenance of multiple representations (numerical and linguistic) that support pre-algebraic thinking (e.g., Kieran, 1992; Sfard & Linchevski, 1994).

Only a handful of relevant studies, have investigated the connection between arithmetic (calculations or word problems) and algebra. Beyond work by the Carpenter group (e.g., Carpenter & Levi, 2000; Jacobs et al., 2007), Lee, Ng, Bull, Pe, and Ho (2011) and Tolar, Lederberg, and Fletcher (2009) found that arithmetic calculations serve as a platform for algebra. But Lee et al. used word problems as the outcome (to serve as a proxy for pre-algebraic knowledge); Tolar et al. focused on college students; and neither study included word problems as a predictor of algebra knowledge. Fuchs et al. (2012) simultaneously considered calculations and word problems as predictors and found they both uniquely predicted third graders’ understanding of the equal sign and variables. This provides support for Pillay et al.’s (l998) third perspective. Yet, to derive causal inferences, research is needed to examine whether intervention in calculations versus word problems improves children’s pre-algebraic knowledge.

Study Overview and Hypotheses

In the present study, we defined *arithmetic calculations* as skill with number combinations (adding/subtracting single-digit operands) and procedural computation (adding/subtracting whole numbers requiring algorithms). We defined *arithmetic word problems* as linguistically presented problem statements, some of which present irrelevant information and/or charts and figures, for which solutions require adding or subtracting 1- or 2-digit numerals. We defined *pre-algebraic knowledge* as understanding of the equal sign (i.e., solving nonstandard equations with one unknown) and the concept of a variable (i.e., completing function tables). In the text that follows, we use the terms *calculations* to denote arithmetic calculations and *word problems* to denote arithmetic word problems.

We assigned second-grade teachers to calculation intervention or word-problem intervention or control group instruction (to control for maturation and the typical school program). Both intervention conditions involved two-tiered responsiveness-to-intervention (RTI), in which (a) whole-class intervention replaced a portion of the classroom teacher's mathematics block for all students in the class, and (b) adult tutors provided supplementary intervention in small groups to children at risk for poor outcomes (i.e., children with the lowest pretest mathematics scores). This small-group tutoring supplemented whole-class intervention. We contrasted the efficacy of calculation and word-problem intervention against each other (and against the business-as-usual control group) on calculation, word-problem, and pre-algebraic knowledge outcomes. Calculation intervention provided no instruction on word problems. Word-problem intervention provided no instruction on calculations: Children were directed to use whatever methods their classroom teachers had taught them to calculate answers.

Our major purpose was to examine connections among calculations, word problems, and pre-algebraic knowledge. Our major contrast, therefore, was between the two active RTI conditions: one focused on calculations and the other on word problems, both of which controlled for instructional time on the relevant domain. At the same time, including a business-as-usual control group controlled for maturation and history effects and permitted conclusions about whether students in one or more of the RTI conditions made more progress than would have occurred without RTI. Our hypothesis was that effects are specific to the focus of intervention. We expected calculation intervention to result in superior calculation outcomes, compared to word-problem intervention and to control. We expected word-problem intervention to produce superior word-problem outcomes, compared to calculation intervention and to control.

Our second hypothesis was that both forms of arithmetic intervention (calculation intervention and word-problem intervention) improve pre-algebraic thinking, such that both conditions would outperform the control group. But we expected word-problem intervention to produce stronger pre-algebraic knowledge than calculation intervention, based on the assumption that arithmetic word problems involve greater symbolic complexity than arithmetic calculations and rely more on the type of mental flexibility, manipulation of symbolic associations, and maintenance of multiple representations that support pre-algebraic thinking. (See "Frameworks" in method section for linkages between calculation intervention and pre-algebraic thinking and between word-problem intervention and pre-algebraic thinking.) Such findings would provide strong evidence that calculation and word-problem skill are separable; that development of either form of arithmetic skill provides a route toward pre-algebraic knowledge, when instruction is designed to provide linkage; but that connections with algebra are stronger for word problems than calculations, with development of word-problem skill providing a stronger route to pre-algebraic competence. We included outcomes that were proximal to and distal from intervention. *Proximal outcomes* refer to measures aligned with at least one of the two active treatment conditions. *Distal outcomes* refer to measures that extend beyond alignment with one or more of the two active treatment conditions and tap students' ability to generalize beyond problem types they learned within intervention.

A secondary purpose of this study was to assess the efficacy of (a) the second-grade two-tiered calculation intervention, based in part on previously validated whole-class (Fuchs et al., 1997) and tutoring (Fuchs et al., 2009, 2011) programs at third grade, and (b) the second-grade two-tiered word-problem intervention, based in part on previously validated whole-class (Fuchs, Fuchs, Craddock, et al., 2008) and tutoring (Fuchs et al., 2009, 2011) programs at third grade. These programs are referred to, respectively, as *Math Wise* and *Pirate Math*. We explain the conceptual framework and methods for calculation intervention, for word-problem intervention, and for instructional linkages between each form of intervention and pre-algebraic knowledge in the method section and in the online supplementary method section.

Method

Participants

Selection—Participants were selected from 1917 children with consent in 127 second-grade classrooms taught by 96 teachers in 25 schools in a metropolitan school district across four cohorts (one per year for 4 years). Some teachers participated in more than one study cohort, hence the discrepancy between number of classrooms and teachers. Selection of the sample occurred in three steps. First, we conducted whole-class screening with these 1917 children on calculation and word-problem measures (see Measures) to create three achievement strata: low on both domains; low in one domain; and low on neither domain (cut points were based on relations between the screening measures and year-end outcomes in a pilot study with a similar population; Fuchs, Zumeta, et al., 2010).

In the second step, to select students for pre/posting, we randomly sampled students stratifying by the three achievement strata at the start of each school year; the goal was to represent students in each stratum but sample more students from lower strata, given the study's focus on RTI. In the third step, we administered the 2-subtest Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) to exclude 92 children scoring <9th percentile on both subtests (because the study was not about intellectual disability). As a result, 1327 of 1917 students were screened into the study. Of these students, 108 completed none or only a small portion of pretesting: 106 moved after screening but before pretesting was completed (the district experiences a disproportionate amount of moving in the first 6 weeks of school) and two students' special education schedule precluded participation. Thus, 1219 students were pretested: 13–123 students per school; 4–47 per teacher (as high as 47 due to multiple cohorts/classrooms for some teachers); and 4–14 per classroom.

Study condition assignment—Stratifying by school, teachers (and their classrooms) were randomly assigned to three conditions: ~37.5% to calculation intervention (33 teachers; 435 students); ~37.5% to word-problem intervention (35 teachers; 459 students); and ~25% to control (28 teachers; 325 students). This maximized research-principled intervention, while maintaining a large enough control group. Once assigned, teachers who participated across cohorts remained in their condition, with ~25% of classrooms in each condition having teachers in multiple cohorts. (In analyses, we nested students of the same teacher across cohorts.)

Although this study was designed as a cluster randomized design, it was carried out as a quasi-experiment for two reasons. First, the study occurred in some schools where teachers had previously been randomly assigned to calculation intervention (CAL) or word-problem intervention (WP) as part of a pilot study. Because the present study relied on the same interventions, we deemed it necessary to maintain teachers in their originally assigned conditions. This involved 12 CAL teachers, 12 WP teachers, and six control teachers. Second, two CAL, three WP, and four control teachers were directed to implement a condition that differed from their assigned condition, and two teachers represented a combination of the above. In these 11 cases, the assignment error was due to researcher misreading of the randomization sequence; in no case was condition determined according to teacher, school, or researcher preference. Below, we present analyses based on the full sample. However, given these issues, we conducted supplemental analyses on the subset of teachers/classrooms whose assignment was according to design (n = 55 teachers). For these analyses, which are available in a supplementary online file (LINK), results were substantively similar: The conditions were always ordered in the same way, and treatment was always either significant or not as reported in the full analyses. Minor differences are described more fully in the online supplementary file. So treatments did not impact student and classroom performance differentially as a function of how assignment occurred. (Note that in the supplementary file, we only report effect sizes in terms of Cohen d or Hedges g statistic, since that is the primary effect size metric; the other effect sizes metrics included in this paper were also similar.)

Attrition—Of the 1219 students, 117 (9.5%) moved after pretesting but before posttesting. This attrition was comparable across conditions ($p > .05$; 11.26% CAL, 8.28% WP, 9.23% control), and on pretest data, students who left did not differ from those who remained. The analyzable sample thus comprised 1102 students from 25 schools, 96 teachers, and 127 classrooms. Across four cohorts, 24% of students were low on both domains; 14% were low on one domain; and 62% were low on neither domain; 51% were female; 83% received subsidized lunch; 13% were English as Second Language; 42% were African American, 27% were white, 23% were Hispanic, and 7% were other; and 5% received special education services. Students' mean age was 7.55 years (SD = 0.39). In line with the study design, 386 students were in CAL (35.0%), 421 in WP (38.2%), and 295 in control (26.8%). The number of students in Cohorts 1–4, respectively, was 310, 292, 267, and 233. Table 1 shows raw score means and standard deviations (*SD*s) for each screening measure and standard scores means and *SD*s for nationally-normed screening measures, as a function of condition. There was no significant difference as a function of treatment condition on any screening measure.

Measures

Screening—From the *Second-Grade Calculations Battery* (SGCB; Fuchs, Hamlett, & Powell, 2003), we administered four subtests of single-digit addition and subtraction in groups: Sums to 12, Sums to 18, Minuends to 12, and Minuends to 18. For each, students have 1 min to complete 25 problems (alpha on this sample: .85 to .93). The screener was Sums to 12. The other three subtests indexed calculation outcomes (see below). *Story Problems* (Jordan & Hanich, 2000) comprises 14 combine, compare, and change word problems, requiring single-digit addition or subtraction. The tester reads each item; children follow along on paper (alpha on this sample: .87). *WASI* (Wechsler, 1999) is a 2-subtest individually administered measure of general cognitive ability (reliability = .92).

Calculations—We used two outcomes, a composite of experimental measures of proximal effects and a composite of commercial measures of distal effects, which correlated .53. The proximal effects composite included five group-administered *SGCB* subtests (Fuchs et al., 2003): Sums to 18, Minuends to 12, and Minuends to 18 (see above) and 2-Digit Addition and 2-Digit Subtraction. For each 2-digit subtest, students have 3 min to complete up to 20 problems (alpha on this sample: .96 and .87). A single-factor solution across these subtests provided a good fit to the posttest data.[2]

The distal effects composite included *Wide Range Achievement Test-3-Arithmetic* (WRAT; Wilkinson, 1993), *KeyMath-Revised* (KM; Connolly, 1998) *Addition*, and *KM Subtraction*. WRAT was administered in groups; KM individually. Each measure progresses from 1-digit addition and subtraction to 2-digit addition and subtraction to whole-number multiplication and division to problems involving fractions, decimals, and more complex calculation skills. On WRAT, children who do not meet a basal on the written calculation items also complete individually-administered counting and symbolic comparison items. Alpha on this sample for the three measures, respectively, was .93, .84, and .81. A single-factor solution across the three subtests provided a good fit to the posttest data.

Word problems—We used two outcomes, an experimental measure of proximal effects and a factor score of commercial measures of distal effects, which correlated .52. The proximal measure was *Second-Grade Story Problems* (Fuchs et al., 2009), which includes 18 problems (never used for instruction) representing combine, compare, change problem types, with missing information in all three positions of the problem schema, with and without irrelevant information, charts, or graphs. Solutions require 1-digit addition and subtraction. In groups, the tester reads a problem aloud; students follow along on paper and have 1 min to write a constructed response (i.e., not select a response from a set of choices) before the tester reads the next problem. Each problem is scored for correct math (1 point) and label (1 point) to reflect processing of the problem statement and understanding of the problem's theme. Alpha on this sample was .88.

The distal effects factor included *KM-Revised Problem Solving* (Connolly, 1998) and *Iowa Test of Basic Skills-Data Interpretation and Problem Solving* (IOWA; Hoover, Hieronymous, Dunbar, & Frisbie, 1993). KM-Problem Solving includes 18 word problems of increasing difficulty, which involve all four operations representing routine word problems with transparent solution strategies; non-routine word problems without clear

[2]Factors were created for proximal computation, distal computation, distal problem solving, and proximal pre-algebraic knowledge, as well as the pretest factors of computation and problem solving. For proximal pre-algebra, distal problem solving, and the pretest problem solving covariate, there were only two measures, so a multiple factor solution is not useful. To compute factors, we used the principal factors method, with the squared multiple correlation on the diagonal initially, and the fit of a single factor was good. The computational pretest factor had only one eigenvalue above 1, and even the second factor did not overlap with the proximal-distal distinction we made. Similarly, for proximal computation outcomes, the first eigenvalue was 3.01, whereas the next eigenvalue was .31. A similar pattern was evidenced for distal computation. Perhaps more importantly, we chose the measures to conceptually represent the six types of outcomes, and based our hypotheses on those types of outcomes.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

solution strategies; and items requiring students to demonstrate comprehension of a word problem without solving it. Administration is individual; items are read aloud; responses are constructed. Testing is discontinued after three consecutive errors. Alpha on this sample was .74. IOWA is administered in groups. It includes 22 word problems representing taught and untaught problem types; numbers in tables and graphs are required to solve some items. Responding is in multiple-choice format. Alpha on this sample was .81.

Pre-algebraic knowledge—Pre-algebraic knowledge outcomes, which were administered only at posttest, included proximal and distal measures. They correlated .50. The proximal measures were *Find X* and *Number Sentences*. Each is administered in groups; requires constructed responses; and begins with the tester modeling a sample problem. With Find X (Fuchs et al., 2009), students solve standard equations (a + b = c or d – e = f) that vary the position of the unknown occurring in all 3 positions. Alpha on this sample was .91. With Number Sentences (Fuchs et al., 2009), the tester reads eight word problems aloud; students have 30 sec to write the standard mathematical equation representing the problem structure (students do not find solutions), with the unknown again occurring in all 3 positions. The score is the number of correct equations. Alpha on this sample was .84. These measures were combined into a factor score. (Note that we did not include items on these proximal measures with operations on the right side of the equal sign or with operations on both sides of the equal sign. This is because such word-problem types were not taught as part of the word-problem intervention and, as such, those items would not have qualified as proximal to the word-problem intervention. Such items were measured in the distal measures.)

For Cohort 1, the distal measure was *Dynamic Assessment of Algebraic Knowledge* (DA; see Fuchs, Compton, et al., 2008 for details), an individually administered measure of children's responsiveness to instruction on finding the missing variable in addition expressions (e.g., $x + 5 = 11$ or $6 + x = 10$), simple multiplication expressions (e.g., $3x = 9$), and equations with two missing variables (e.g., $x + 2 = y – 1$; $y = 9$) (Skill A, Skill B, and Skill C, respectively). Mastery of each skill is assessed before instructional scaffolding begins and recurs after each level of scaffolding. If mastery occurs, the tester administers a generalization problem (for Skill A, $3 + 6 + x = 11$; for Skill B, $14 = 7x$; for Skill C, $3 + x = y + y$; $y = 2$) and moves to the next skill. If mastery does not occur, the tester provides the first (or next) level of instructional scaffolding, which is followed by the mastery test. Each level of scaffolding increases instructional explicitness and concreteness. If a student fails to achieve mastery after all five scaffolding levels for a given skill, testing is terminated. Scores range from 0–21 (0 = never mastered any skill; 21 = mastered each skill on the pretest and got each bonus problem correct). Alpha on this sample was .84. The outcome was a sample-based *z*-score.

In Cohorts 2–4, a composite was generated across DA and the *Test of Pre-Algebraic Knowledge* (Fuchs, Seethaler, & Powell, 2009), which comprises two types of problems. The first problem type (20 items) involves mathematical equivalence statements with letters standing for missing quantities: 18 in nonstandard format (e.g., $y + 4 = 9 + 3$); two in standard format (i.e., $1 + 5 = x$). The next problem type (4 items) involves function tables, each of which shows a 2-column table. The first column shows a variable; the second shows a function involving that variable; each row shows a value for the variable and the resulting

NIH-PA Author Manuscript

NIH-PA Author Manuscript

value for the function. In one row, the value of the function is empty; the task is to complete that row. The functions are $x + 3$, $y - 6$, $2x + 1$, and $3y$. In groups, the tester demonstrates how to complete a sample problem for each problem type. Students have 8 min to complete the first problem type and as much time as needed (until all but two students finish) to complete the second problem type. The correlation between the two problem types was .54. The pattern of results was the same for the two problem types, so we used the total score. Alpha on this sample was .88. In Cohorts 2–4, the score was a sample-based *z*-score composite of DA and Test of Pre-Algebraic Knowledge.

Study Conditions—(For additional information, see online supplementary method file.)

The study conditions were business-as-usual control, 2-tiered CAL intervention, and 2-tiered WP intervention. Tier 1 was 34 whole-class intervention lessons (2 lessons per week for 17 weeks; 40–45 min per lesson) for all students in the class. Researcher-delivered whole-class instruction substituted for ~185 of ~300 min of classroom teachers' weekly business-as-usual mathematics instruction.

Tier 2 was 39 tutoring lessons (3 times per week for 13 weeks, beginning in Weeks 4–5 of Tier 1 instruction; 2–3 children per group; 25–30 min per lesson) provided to 272 students. The benchmark for low performance to determine eligibility for tutoring was <7 on calculation and word-problem screeners. This yielded 320 students who were eligible for tutoring. In Cohorts 1 and 4, we accommodated more students due to additional resources (as typically done in RTI). So CAL students scoring <7 on calculations but >7 on word problems also were eligible, as were WP students scoring <7 on word problems but >7 on calculations. Also, in each cohort, before finalizing tutoring decisions, we asked teachers to confirm the appropriateness of selections based on classroom observations. With this teacher input, 50 students who were eligible for tutoring did not receive it, and 2 students who were not eligible did receive it. So although 320 students were eligible for tutoring according to the benchmarks we had set, 272 students received tutoring: 84 in Cohort 1 (42 in CAL; 42 in WP); 50 in Cohort 2 (25 in CAL; 25 in WP); 72 in Cohort 3 (34 in CAL; 38 in WP); and 66 in Cohort 4 (34 in CAL; 32 in WP).

Below, we describe (a) the framework for CAL and WP intervention, with linkages in each program to pre-algebraic knowledge; (b) the nature of control group instruction and distinctions between control and the intervention conditions; and (c) fidelity of implementation. Program manuals, which include lesson guides, are available from the first author, under the title *Math Wise* for CAL and *Pirate Math* for WP. See the on-line supplementary method file for more detailed information on (a), (b), and (c) and for information on the structure of whole-class CAL and WP instruction, on the structure of CAL and WP tutoring, and on the research assistant teachers and tutors and how they were prepared and supported.

Framework for CAL intervention—CAL intervention incorporated two major emphases that reflect understanding about how children develop competence with simple arithmetic and procedural calculations (e.g., Fuchs, Geary, et al., 2013; Fuson & Kwon, 1992; Geary et al., 2008; Groen & Resnick, 1977; LeFevre & Morris, 1999; Siegler & Shrager, 1984). The

NIH-PA Author Manuscript

NIH-PA Author Manuscript

first emphasis was *interconnected knowledge about number* (e.g., cardinality, inverse relation between addition and subtraction; commutate property). For example, students used manipulatives to explore how a target number can be partitioned in different ways. They focused on part-whole knowledge with number families, grouping families and using visual displays/blocks to show how/why four problems make a family and to explore the inverse relation between addition and subtraction. The number knowledge emphasis in CAL also had a strong focus on tens concepts and place value. Students practiced counting by 10s with a number list; explored relations between ones and tens and the meaning of zeros in the ones and tens places; used and regrouped manipulatives to represent 1-and 2-digit numbers; and identified smaller and larger numbers using place value and the number list.

The second major emphasis in CAL intervention was *practice*. Students were taught and practiced efficient counting procedures for solving 1-digit problems and 2-digit plus 1-digit problems that do not require regrouping. Practice required students to generate many correct responses to help them form long-term representations to support retrieval. Students were also taught and practiced efficient procedures for identifying when regrouping was required in addition and subtraction problems and for actual regrouping.

CAL intervention was divided into six units: (a) equal sign as a relational term; (b) addition concepts and operational strategies for problems for which retrieval is a viable strategy; (c) concepts and operational strategies for similar problems involving subtraction; (d) concepts and operational strategies for addition problems with regrouping; (e) concepts and operational strategies for subtraction problems with regrouping; and (f) review.

Framework for WP intervention—Our framework for studying word problems was based on Kintsch and colleagues (Cummins, Kintsch, Reusser, & Weimer, 1988; Kintsch & Greeno, 1985; Nathan, Kintsch, & Young, 1992), who pose that word-problem solving is an interaction between problem-solving strategies and language comprehension processes. This model assumes that general features of the text comprehension process apply across stories, informational text, and word-problem statements, but the comprehension strategies, the nature of required knowledge structures, and the form of resulting macrostructures and situation and problem models differ by task. According to this model, memory representations of word problems have three components. The first involves constructing a coherent structure of the text's essential ideas. The second, the situation model, requires supplementing the text with inferences based on the child's world knowledge; this includes informal knowledge about conceptual relations among quantities. The problem solver coordinates this information with the third component – *problem models* or *schema* – to formalize the conceptual relations among quantities. The schema guides application of solution strategies. At second grade, combine, compare, and change problem types are the major schema. The model poses that this process makes strong demands on working memory, reasoning ability, and language comprehension.

In terms of working memory and reasoning ability, consider a combine problem (two parts are combined to make a total): *Joe has 3 marbles. Tom has 5 marbles. Tom also has 2 balls. How many marbles do the boys have in all?* The problem solver processes sentence 1 to identify object is marbles; quantity is 3; actor is Joe; but Joe's role is unknown. This is placed in short-term memory. In sentence 2, propositions are similarly coded and held in memory. In sentence 3, *balls* fails to match the object code in sentences 1 and 2, signaling that 2 balls may be irrelevant; this is added to memory. In the question, the quantitative proposition *how many marbles* and the phrase *in all* cues the problem solver that this problem falls in the combine schema. So the problem solver assigns the role of superset (total) to the question; checks information held in short-term memory to assign subset roles (the two parts); and rejects 2 balls as irrelevant. Filling in these slots of the schema in this way triggers a set of problem-solving strategies. The hope is that with typical school instruction, children gradually construct the combine schema on their own, just as they devise strategies for handling the demands on working memory and reasoning this problem-solving sequence involves.

Our schema-based approach to WP intervention explicitly teaches children the underlying structure of combine, compare, and change schema, using real-life scenarios and role playing with stories that have no unknowns. The teacher (a) transitions from complete stories to problem statements with missing information and (b) introduces graphic representations to formalize the quantitative relations underlying each schema and provide opportunities for students to place known numbers and variables into the graphic representations. The teacher then transitions to “meta-equations” to represent the schema and teaches step-by-step strategies that begin with identifying problem statements as combine, compare, or change schema and then building the propositional text structure. Schema-based instruction facilitates connections among the situation model, schema, and productive solution strategies by making these connections explicit. It also provides children with strategies that reduce demands on working memory and reasoning. The child RUNs through the problem: Reads it, Underlines the question in which the object code (marbles) is revealed, and Names the explicitly taught combine schema. This prompts the child to write the combine meta-equation (P1 + P2 = T for the above problem). The child then re-reads the problem statement. While re-reading, he/she replaces P1 and P2 with quantities for each relevant “part” and crosses out irrelevant objects/numbers. This reduces the burden on working memory and reasoning, as it provides the equation for problem solving and sets up the solution equation.

Word-problem solving also relies on language comprehension processes. As per Kintsch and Greeno (l985), children learn to treat important vocabulary and language constructions in a special, task-specific way, including extensions to ordinary usage for terms (e.g., *all or more*) to more complicated constructions involving sets (*in all* and *more than)*. But for many children, this assumption is shaky. Cummins et al. simulated incorrect problem solving with two types of errors: incorrect math problem-solving processes versus language processing errors. Correct problem representation depended more on language, and changing wording in only minor ways dramatically affected accuracy. As Nathan et al. (1992) concluded, instruction must focus on language processes as well as the mathematical aspects of word-problem solving. Our approach to schema-based instruction differs from other forms of schema-based instruction (e.g., Jitendra, Star, Rodriguez, Lindell, & Someki, 2011; Jitendra et al., 2009), in part, by providing explicit instruction on the language comprehension demands specific to combine, compare, and change problem types. The major challenges we address are (a) underdeveloped representations of relational terminology and constructions (e.g., *more/less than; older; stronger*) for compare problems; (b) discriminating relational vocabulary and constructions from confusable ones (e.g., *Tom has 5 fewer marbles than Jill*, as in compare problems, vs. *Tom had 5 marbles and then he got 2 more*, as in change problems); and (c) under-developed representations of vocabulary related to quantities (e.g., *amount* refers to quantity) and taxonomic relations at superordinate levels (e.g., 2 dogs+3 cats=5 animals; McGregor et al., 2002), which are important for combine problems.

WP intervention was divided into five units: (a) foundational skills for the word-problem content (i.e., equal sign as a relational term; strategies to find x; strategies for checking word-problem work); (b) combine program; (c) compare problems; (d) change problems; and (f) review. The program typically provides explicit conceptual and strategy instruction on 1- and 2- calculations (e.g., Fuchs et al., 2009), but for the present study, we removed all instruction on calculations. When students asked questions or needed corrective feedback on calculations, they were told to use the strategies they learned from their classroom teachers.

Linkages with pre-algebraic knowledge—CAL and WP intervention incorporated instructional linkages to pre-algebraic knowledge, as per Pillay et al. (1989). This occurred in two ways. First, both CAL and WP intervention explicitly focused on *understanding the equal sign as a relational symbol* (Jacobs et al., 2007). Some work (Baroody & Ginsburg, 1983; Blanton & Kaput, 2005) suggests that teachers' consistent use of the phrase *is the same as* (instead of *equals*) with young children is associated with improved understanding of the equal sign. Short-term experiments with intermediate age students show that explicit instruction on the meaning or location of the equal sign can enhance equal sign understanding and performance on open, nonstandard equations (e.g., 6 + 4 + 7 = 6 + __; McNeil & Alibali, 2005; Rittle-Johnson & Alibali, 1999). Powell and Fuchs (2010) showed that third graders with mathematics difficulty who received schema-based tutoring plus equal-sign instruction performed better than students who received schema-based tutoring alone on closed equations and some types of word problems.

Second, as discussed, WP intervention taught children to represent the underlying structure of schemas in terms of "meta-equations": for combine problems, P1 + P2 = T; for compare problems, Bigger minus Smaller = Difference (B – s = D); for change problems, Start plus/minus Change = End (St +/− C = E). Children were taught to identify the problem type and write the corresponding meta-equation; re-read while replacing slots in the meta-equation with information from the problem statement (including x for the unknown); and solve for x (x could occur in any of the three slots of the equation). This has been shown to encourage pre-algebraic thinking in second graders (Fuchs, Zumeta, et al., 2010). Because WP intervention provided this additional linkage with algebraic thinking over CAL and because WPs may involve greater symbolic complexity than calculations (as outlined in the introduction), we expected WP intervention to stronger pre-algebraic knowledge than CAL.

Distinctions between control the CAL/WP intervention—Classroom teachers relied primarily on the basal program *Houghton Mifflin Math* (Greenes et al., 2005) to guide mathematics instruction. Their curricular content aligned with the content in CAL intervention (1- and 2-digit adding and subtracting) and WP intervention (combine, compare, and change word problems). In this way, control students received calculation and word-problem instruction relevant to the study. The amount of whole-class instruction was comparable in all three conditions, but tutored children in CAL and WP intervention received more instruction than some of the children in control group who would have been eligible for tutoring (instruction was of similar time for control group students who participated in the school's intervention period in math). Results, however, indicated no interaction between tutoring eligibility status and treatment condition.

Based on analysis of *Houghton Mifflin Math* (Greenes et al., 2005) and teacher reports, key distinctions between the control and CAL conditions were as follows. (1) Control group instruction did not address the equal sign as a relational term. (2) Control group instruction focused less on number knowledge and more on procedures. (3) CAL provided greater emphasis on development of fluency with efficient counting strategies. Important commonalities between the control and CAL conditions were as follows: instruction addressed one problem type at a time; provided students with explicit steps for deriving solutions; and relied on worked examples, guided group practice, and independent practice with corrective feedback.

Key distinctions between the control and WP conditions were as follows. (1) Control group instruction did not address the equal sign as a relational term. (2) Control group instruction emphasized a metacognitive approach to solving word problems, in which students used guided generic questions (not specific to problem types) to plan, solve, and reflect on the content of word problems; WP did not employ this general set of metacognitive strategies. (3) In contrast to WP, there was no attempt in the control condition to explicitly teach students to understand word problems in terms of the combine, compare, or change schemas. (4) Control group instruction provided strong emphasis on computational requirements for problem solution; WP intervention provided none. (5) Although control group instruction allocated 3 weeks of instruction to finding missing addends, it focused substantially less on this topic and did not connect this topic to the structure of word problems. (6) Control group instruction taught children to rely on keywords (e.g., *more* is a signal to add the numbers in the problem), a common approach in schools; WP avoided keywords (because they only produce correct solutions ~50% of the time). Important commonalities between the control and WP conditions were as follows: instruction addressed one problem type at a time; focused on concepts underlying the problem type; provided students with explicit steps for deriving solutions; and relied on worked examples, guided group practice, and independent practice with corrective feedback.

Fidelity—Prior to the first whole-class and tutoring session, research staff agreed on the essential information in each lesson and made a checklist of points for each lesson. This was done for CAL and WP whole-class instruction and tutoring. Each session was audiotaped. At the study's end, RAs independently listened to a random sample of tapes while completing checklists to identify the percentage of points addressed. We sampled 20% of whole-class instruction tapes equitably within conditions, RA-teachers, classrooms, and lesson types; we sampled 20% of tutoring tapes equitably within conditions, RA-tutors, tutoring groups, and lesson types. For whole-class intervention, the mean percentage of points addressed was 95.87 (SD = 1.40) for CAL and 94.86 (SD = 1.85) for WP, t(16) =

1.70, $p = .110$. For tutoring, the mean percentage of points addressed was 96.06 ($SD = 2.83$) for CAL and 96.34 ($SD = 3.28$) for WP, $t(35) = 0.58$, $p = .563$. (In these fidelity analyses, teacher/tutor was the unit of analysis; *t* tests are for dependent samples because each teacher/tutor taught in both conditions.)

Procedure

Testing occurred in classrooms for measures administered in groups or other school locations for measures administered individually. Students were assessed on screening measures in September; on pretest measures in October. Research assistants (RAs) delivered whole-class instruction in November-March and tutoring in December-March. Posttesting occurred in March. Testers were trained to criterion on each measure. All individual test sessions were audiotaped, and a random sample of sessions was rescored from tapes by a second RA, with agreement of 98%. All data were double entered/verified.

Analysis Plan

The first step was an unconditional model, to evaluate the effect of clustering at the teacher level (i.e., one-way ANOVA with random effects; Luke, 2002; Sullivan et al., 1999). Our models also included school as a third level; the purpose here was to account for clustering at the school level (rather than expecting treatments to operate differentially at the school level). Two intraclass coefficients (ICCs) were denoted for (a) teacher (within school), where the numerator was variance due to that clustering, and for (b) school, where the numerator was variance due to school clustering; the denominator in both cases are those terms plus residual variance (i.e., total variance).

The second step added pretest as a student-level fixed effect to explain posttest variability (i.e., one-way ANCOVA with random effects). The key parameter was the value of the regression effect of the pretest on posttest in a given classroom in a given school (these regression effects were not set to vary across classes/schools). The pretest covariate(s) included student and teacher level measures of (a) factor score across the WRAT and the five SGCB measures for calculation outcomes, or (b) a factor score from Second-Grade Story Problems and KM-Problem Solving for word-problem outcomes (IOWA was not administered at pretest). Both of these pretests were used for pre-algebraic knowledge outcomes. Also in this second step, though after pretest, we added additional student-level covariates, which in this study were sex, ethnicity, reduced/free lunch (RFL) status, English as Second Language (ESL), and tutoring eligibility status.

The third step was to add the treatment effect into the fixed portion of the model, at the level of the teacher. A key difference from the student-predictor model is that the intercept of this equation is now conditional on the treatment effect (its regression effect). Other teacher level predictors were then added at this third step. Specifically, we also added pretest as a teacher-level fixed effect in these initial models to help alleviate issues associated with a quasi-experiment. We also added cohort, since new teachers were assigned to conditions for each cohort. In each case, the teacher-level fixed effects were never contributory over student level pretests, and so were dropped from remaining models. Finally at this step, we also included interactions between treatment and cohort, which are described below if they occurred.

The fourth step moved from the fixed portion of the model to the random portion. Students were at level 1, teachers at level 2, and schools at level 3. In addition to a random intercept (denoting variability in outcomes), we also evaluated whether a random effect of student-level pretest could be added (e.g., denoting variability of slopes – which would indicate that the relation between pre- and posttest varied across teachers). These additional random effects were not contributory to any models and are not discussed further. That is, there were no systematically differential relationships of pre- to posttest across teacher/classroom. Finally, given the relatively small number of school units, we also considered the role of school as a fixed, rather than random effect; in doing so, school never interacted with treatment, nor did it alter the treatment effects substantively.

Model fit was estimated with restricted maximum likelihood (REML) in PROC MIXED. Fit statistics included evaluation of deviance (−2 log likelihood) and other means (e.g., Akaike Information Criteria or AIC; Bayesian Information Criteria or BIC). Denominator degrees of freedom were computed according the Kenward-Rogers algorithm. We computed three types of effect sizes (ESs): a measure analogous to R^2, a proportional reduction in variance (PRV), and a measure analogous to a more traditional Cohen *d* or Hedges *g* statistic. First, at a global level, we outputted predicted values from the final model and correlated these with obtained values; then we squared the result (Peugh, 2010; Singer & Willet, 2003). The resulting value, however, does not distinguish the contributions of the individual parameters or levels. Second, at a local level, we computed a proportional reduction in variance (PRV; Peugh, 2010; Raudenbush & Byrk, 2002; Singer & Willett, 2003) that is focused on a specific model parameter (treatment) and level (school/teacher). We subtracted the variance components of the final model, not from an unconditional model but rather from a model with all other predictors except treatment, thus isolating treatment's specific added contribution. The PRV effects are only relative values (comparing one model to another); they are not an absolute amount of variance explained (Ma et al., 2008). Third, we computed ESs for treatment conditions by subtracting estimates between pairs of treatment condition and dividing by the square root of the student level residual variance (and thus is interpretable as a mean difference divided by a standard deviation, similar to a *d* or *g*), when treatment was the only fixed effect in the model. The square root of the student residual variance was used in the denominator rather than the raw *SD*, because the former considers the clustering effect (Tymms, 2004). We present each of these indices (referred to as R^2, PRV, and ES, respectively, in the text that follows) to assist readers in interpreting the relative effects of treatment in this study.

Results

Table 1 shows raw score means and *SD*s for each outcome measure and standard scores means and *SD*s for nationally-normed outcome measures. There was no significant difference as a function of treatment condition on any pretest measure. Tables 2 to 4 show results for the calculations, word-problems, and pre-algebraic knowledge, respectively. Proximal and distal outcomes are shown in the same table. The top portion of each table shows results of the unconditional models. The bottom shows results for the full/final model, including effects for treatment, cohort, tutoring eligibility status, pretest performance covariates, and demographic covariates pertinent to the outcome. Covariates and tutoring eligibility are level 1 (L1, student) predictors; cohort and treatment are level 2 (L2, teacher) predictors.

Proximal calculation effects (Table 2, left side)—For the unconditional model, the ICCs (computed according to Equations 1b and 1c) were .12 for teacher and an additional .07 for school (i.e., 12% of the variance in proximal computations was accounted for by clustering of students at the teacher level, with an additional 7% at the school level). Model fit statistics were as follows: deviance = 2829.2, parameters = 3, AIC = 2835.9, BIC = 2839.5. When L1 predictors were added to the model, ESL status, RFL status, sex, and ethnicity were not significant and so were dropped. When L2 predictors (treatment, cohort) were added, there was no interaction and no effect for tutoring eligibility status. In the final model (R^2 = .49), there were significant effects for pretest ($p < .001$; higher outcomes with stronger pretest performance), cohort, $F(3,71.9) = 3.85$, $p < .014$, and treatment condition, $F(2,92.8) = 12.97$, $p < .001$. Follow up to the treatment effect indicated CAL outperformed WP ($p < .001$; ES = 0.41) and control ($p < .001$; ES = 0.55), but the latter two groups did not differ ($p = .223$; ES = 0.14). Model fit statistics for the final model were: deviance = 2277.2, parameters = 3, AIC = 2283.2, BIC = 2286.8. The PRV at the level of teacher and school due to treatment condition (relative to a model with all predictors except treatment) was .21 (i.e., adding treatment reduced variance between teachers/schools by 21% relative to the previous model).

Distal calculation effects (Table 2, right side)—The unconditional model showed an ICC of .07 for teacher and .06 for school. Model fit statistics were as follows: deviance = 2693.0, parameters = 3, AIC = 2699.0, BIC = 2702.7. For the model adding L1 predictors, ESL status was not significant and so was dropped. When L2 predictors (treatment, cohort) were added, there was no interaction. In the final model (R^2 = .38), there were significant effects for pretest ($p < .001$), sex ($p < .035$), ethnicity ($p < .001$), RFL status ($p < .001$), tutoring eligibility status ($p < .004$), and treatment condition, $F(2,93.9) = 9.96$, $p < .001$, but not cohort ($F[3,56.5] < 1$). Across treatment conditions and controlling for all other variables in the model, outcomes were stronger for students with higher pretest scores, for those not receiving RFL, for those not eligible for tutoring, for girls, and for children who were not African American. Follow up to the treatment effect indicated CAL outperformed WP ($p < .003$; ES = 0.30) and control ($p < .001$; ES = 0.38), and the latter two groups did not differ ($p = .116$; ES = 0.07). Model fit statistics for the full/final model were: deviance = 2283.1, parameters = 3, AIC = 2289.1, BIC = 2292.7. Relative PRV for the addition of treatment condition was .22.

Proximal word-problem effects (Table 3, left side)—For the unconditional model, the ICC was .25 for teacher and .01 for school. Model fit statistics were as follows: deviance = 7385.0, parameters = 3, AIC = 7391.0, BIC = 7394.6. When L1 predictors were added, ESL status, RFL status, and ethnicity were not significant and so were dropped. The final model (R^2 = .52) showed significant effects for pretest ($p < .001$), sex ($p < .003$), tutoring eligibility status ($p < .021$), cohort $F(3,43) = 6.43$, $p < .001$, and treatment condition, $F(2,99.7) = 119.07$, $p < .001$. Across treatment conditions and controlling for all other variables in the model, outcomes were stronger for students with higher pretest scores, for girls, and for students not eligible for tutoring. Follow up to the treatment effect indicated WP outperformed CAL ($p < .001$; ES = 1.00) and control ($p < .001$; ES = 0.92), but the latter two groups did not differ ($p = .718$; ES = −0.07). The overall cohort effect indicated Cohorts 3 and 4 outperformed Cohorts 1 and 2. There was also a significant interaction of treatment and cohort, $F(6,157) = 11.29$, $p < .001$, which suggested relative differences between WP and the other groups as a function of cohort; however, WP consistently outperformed the other two groups in each cohort. Model fit statistics were: deviance = 6671.9, parameters = 3, AIC = 6676.9, BIC = 6681.6. Relative PRV for treatment condition was .89.

Distal word-problem effects (Table 3, right side)—For distal word problems, the unconditional model showed an ICC that was less than .01 for teacher and .10 for school. Model fit statistics were as follows: deviance = 2539.7, parameters = 3, AIC = 2545.7, BIC = 2549.3. All considered L1 predictors were retained. The final model ($R^2 = .61$) showed significant effects for pretest ($p < .001$), sex ($p < .021$); tutoring eligibility status ($p < .001$), RFL status ($p < .012$), ESL status ($p < .001$), ethnicity, $F(3,1013) = 10.05$, $p < .001$), but not cohort, $F(3,75.5) = 1.99$, $p = .123$. Across treatment conditions and controlling for all other variables in the model, outcomes were stronger for students with higher pretest scores, for boys, for students not eligible for tutoring, for students receiving RFL or ESL, and for students were not African American. The effect for treatment condition was not significant, $F(2,1045) = 1.56$, $p = .211$. Model fit statistics for the final model were: deviance = 1659.8, parameters = 2, AIC = 1663.8, BIC = 1666.3. Relative PRV for treatment was < .01, and all ES differences were negligible (|range ES| = .01 to .10).

Proximal pre-algebraic knowledge effects (Table 4, left side)—In the unconditional model, the ICC was .35 for teacher and essentially zero for school. Model fit statistics were as follows: deviance = 1898.8, parameters = 2, AIC = 1902.8, BIC = 1905.2. All covariates and predictors were retained except sex, which was not significant. The final model (including L2 predictors, $R^2 = .51$) showed significant effects for both pretest measures (both $p < .001$), RFL status ($p = .045$), ESL status ($p = .014$), ethnicity, $F(3,879) = 5.33$, $p = .002$), and treatment, $F(2,82.1) = 162.86$, $p < .001$. Tutoring eligibility status and cohort were not significant (both $p > .05$). Across treatment conditions and controlling for other variables in the model, outcomes were stronger for students with higher pretest scores and for students not receiving RFL or ESL; outcomes were lower for African American compared to other subgroups and higher for Hispanic than Caucasian. Follow up to the treatment effect indicated WP outperformed CAL ($p < .001$; ES = 1.32) and control ($p < .001$; ES = 1.36), but the latter two groups did not differ ($p = .748$; ES = 0.04). There was a significant interaction of treatment and cohort, $F(6,134) = 3.94$, $p < .002$, suggesting relative differences between WP and the other groups as a function of cohort; however, WP consistently outperformed the other two groups in each cohort. Model fit statistics were: deviance = 1444.0, parameters = 3, AIC = 1450.0, BIC = 1453.7. Relative PRV for treatment condition was .90.

Distal pre-algebraic knowledge effects (Table 4, right side)—The unconditional model showed an ICC of .02 for teacher and .04 for school. Model fit statistics were as follows: deviance = 2893.1, parameters = 3, AIC = 2899.1, BIC =2902.8. For the model adding L1 predictors, RFL status was significant so it was dropped. When L2 predictors (treatment, cohort) were added, there was no interaction. In the final model (R^2 = .48), there were significant effects for both pretests (both $p < .001$), sex ($p < .021$), ethnicity ($p < .001$), ESL status ($p < .030$), tutoring eligibility status ($p < .031$), and cohort ($F[3,71.6] = 4.58$, $p < .006$). Across treatment conditions and controlling for all variables in the model, outcomes were stronger for students with higher pretest scores, for boys, for students receiving RFS or ESL, for students not eligible for tutoring, and for Cohorts 2 and 3 compared to Cohort 1; outcomes were lower for African American compared to other subgroups and higher for Hispanic than Caucasian. There was also a significant treatment effect, $F(2,105) = 3.48$, $p < .035$. WP outperformed CAL ($p < .036$; ES = 0.21) and control ($p < .024$; ES = 0.22); the latter two groups did not differ ($p = .667$; ES = 0.01). Model fit statistics were: deviance = 2183.3, parameters = 3, AIC =2189.3, BIC = 2192.9. Relative PRV for treatment condition was .13.[3]

Discussion

The focus of the present study was connections among three aspects of mathematical cognition in second-grade children: calculations, word problems, and pre-algebraic knowledge. We extended the literature, which is dominated by correlational work, by examining whether intervention conducted on calculations or word problems contributes to improved performance in the other domain and whether intervention in either or both domains contributes to pre-algebraic knowledge. This is the first study to estimate the specificity of word-problem intervention and the first to investigate the specificity of calculation intervention in the same study. It is also the first to assess whether intervention on calculations versus word problems makes a stronger contribution to children's pre-algebraic knowledge. This study design provides a stronger basis than previous, correlational research for determining whether these aspects of mathematical cognition are shared or distinct and for informing practice.

Specificity of Effects of Calculation and Word-Problem Intervention

In terms of the specificity of effects of calculation and word-problem instruction, results were in line with our hypotheses. Intervention improved performance in the targeted domain, but not the other domain. So students who received calculation intervention completed the study with stronger calculation skill than students in word-problem intervention (ES = 0.41 and 0.30 for proximal and distal outcomes, respectively). They also finished the study with stronger calculation skill than students in the business-as-usual control group (ES = 0.55 and 0.38). But on word-problem outcomes, students who received calculation intervention performed comparably to students in the business-as-usual control

[3]We also ran an exploratory analysis to examine whether the pattern of effects of the distal pre-algebra outcomes on the Dynamic Assessment's Skill C (assessing students' ability to learn to solve equations most similar to algebra, e.g., $x + 2=y - 1$; $y=9$) were parallel to the overall composite. Analysis of variance revealed a significant effect, $F(2,1099) = 5.12$, $p = .006$. Follow-up tests revealed that the performance of the WP condition was stronger than that of the CAL group (ES = 0.22).

group (ES = 0.14 and 0.07). On the one hand, this lack of transfer corroborates a handful of prior studies (Fuchs et al., 2009, 2010). It also may not seem surprising, because calculation intervention did not provide students opportunity to practice the word-problem skill their classroom teachers were addressing. On the other hand, because calculation skill is required for word-problem success, one might expect the superior calculation skill of students in the calculation intervention condition to provide them a boost on word problems, especially compared to the control group. We did not find evidence to support this.

Results occurred in parallel fashion for word-problem intervention on the proximal outcome. (We discuss distal outcomes, for which effects were not significant, later.) Students who received word-problem intervention completed the study with stronger skill than students in calculation intervention (ES = 1.00) and stronger performance than students in the business-as-usual control group (ES = 0.92). But on calculation outcomes, students who received word-problem intervention performed comparably to students in the business-as-usual control group (ES = 0.14). This is notable for the following reasons. This is the first study to examine transfer from word-problem intervention to calculation skill (most word-problem intervention studies address the calculation skills required for successful word-problem performance within the intervention or do not assess calculation outcomes). The second reason this lack of transfer from word-problem intervention to calculation skill is notable is that word-problem intervention provided students with many opportunities in every session to apply the calculation skills they were learning from their classroom teachers. Students were also required to correct errors (using whatever procedures their teachers had taught them). Yet, we saw no evidence of transfer.

These findings, in which calculation intervention improved calculation but not word-problem outcomes and in which word-problem intervention improved word-problem but not calculation outcomes, indicate these aspects of mathematical cognition are separable. This supports our hypotheses, which were based on correlational research demonstrating the cognitive abilities underlying word problems and calculations differ (e.g., Fuchs, Fuchs, Stuebing, et al., 2008; Fuchs et al., 2006; Fuchs, Geary, et al., 2010a, b; Geary et al., 2012; Swanson, 2006). Present findings, together with correlational findings, argue for conceptualizing these two aspects of mathematical cognition distinctly. Practically, it suggests that screening for mathematics difficulty requires different measures. It also suggests that standards-setting committees, curriculum designers, and teachers should address these two aspects of mathematical cognition explicitly and deliberately.

By contrast, at the present time, research and practice focus disproportionately on calculations over word problems, perhaps with the assumption that understanding about calculation development pertains to word problems and that instruction on calculations will transfer to word problems. This is unfortunate given the present study’s findings and due to the importance of word problems. Word-problem skill is the best school-age predictor of employment and wages in adulthood (Bynner, 1997; Every Child a Chance Trust, 2009; Parsons & Bynner, 1997; Rivera-Batiz, 1992). Also, word problems can be a persistent deficit even when calculation skill is adequate (Swanson, Jerman, & Zheng, 2008). And the cognitive processes involved in word problems are more numerous than those underlying calculation skill (e.g., Fuchs, 2006, 2010a, b; Fuchs, Fuchs, Stuebing et al., 2008; Geary et al., 2012), which suggests that word-problem difficulty may be more complicated to prevent and remediate.

The Role of Arithmetic Calculations and Word Problems in Pre-Algebraic Thinking

We also extended the literature by examining whether intervention conducted on calculations or word problems and designed to provide linkages between the arithmetic and the pre-algebraic stages of development (Pillay et al., 1998) contributes to pre-algebraic knowledge. Toward this end, calculation and word-problem intervention incorporated a strong focus on relational understanding of the equal sign (Blanton & Kaput, 2005; McNeil & Alibali, 2005; Powell & Fuchs, 2010; Rittle-Johnson & Alibali, 1999). Word-problem intervention also was designed to encourage pre-algebraic insight by teaching children to rely on “meta-equations” to represent the underlying structure of word problems; replace slots of the meta-equation with unknowns and knowns from the problem statement; use x to represent the unknown; and solve for x (Fuchs, Zumeta, et al., 2010).

The goal was to increase understanding about whether arithmetic supports the development of algebra and, if so, whether calculations or word problems contribute in differential ways. Lee et al. (2011) and Tolar et al. (2009) found that calculations serve as a platform for algebra. But Lee et al.’s pre-algebraic knowledge outcome was word problems; Tolar et al. focused on college students; and neither study examined word problems as a predictor of algebra. Fuchs et al. (2012) simultaneously considered calculations and word problems and found that both uniquely predicted third graders’ understanding of the equal sign and variables. We therefore hypothesized that either form of arithmetic intervention (on calculations or on word problems) improves pre-algebraic thinking, such that each intervention condition would outperform the control group. Also, based on the assumption that word problems involve greater symbolic complexity than calculations and rely more on the type of mental flexibility, manipulation of symbolic associations, and maintenance of multiple representations that support pre-algebraic thinking, we expected word-problem intervention to produce stronger pre-algebraic knowledge than calculation intervention. Findings partially supported these hypotheses.

On pre-algebra outcomes, word-problem intervention resulted in superior performance compared to the business-as-usual control group and compared to the calculation intervention condition. This is notable because significant effects occurred not only on pre-algebraic outcomes proximal to word-problem intervention (respective ESs = 1.36 and 1.32). Significant effects also occurred on the distal composite (respective ESs = 0.22 and 0.21), which assessed (a) performance on nonstandard equations and functions that were not addressed in intervention and (b) responsiveness to instruction, via dynamic assessment, problem types also not addressed during intervention: simple multiplication expressions (e.g., $3x = 9$) and equations with two missing variables (e.g., $x + 2 = y - 1$; $y = 9$). Moreover, exploratory analysis was used specifically to examine whether the pattern of effects of the distal pre-algebra outcomes on the Dynamic Assessment’s Skill C (assessing students’ ability to learn to solve equations most similar to algebra, e.g., $x + 2 = y - 1$; $y = 9$) were parallel to the overall composite. These results also revealed a significant effect: The

NIH-PA Author Manuscript

NIH-PA Author Manuscript

performance of the word-problem condition was stronger than that of the calculation intervention group, with an almost identical ES of 0.22.

At the same time, however, contrary to our hypothesis, calculation intervention did not produce superior pre-algebraic performance compared to the control group (ES= 0.04 and 0.01). This is in line with McNeil (2008), who found that children do not benefit much from equal sign instruction when it is given in the context of typical symbolic arithmetic problems. Contrary to Pillay et al.'s (l998) framework, this suggests that calculation intervention, even when it includes a strong focus on relational interpretation of the equal sign, does not contribute to pre-algebraic thinking, relative to control group instruction. However, in line with Pillay et al., results suggest that early arithmetic word-problem intervention helps children make the transition from arithmetic to the pre-algebraic stage of development.

Future research should continue to pursue this issue, given that (a) few studies have examined how students develop competence with algebra; (b) algebra is important for successful participation in the U.S. workforce and to higher forms of learning in mathematics, science, technology, and engineering (NMAP, 2008; RAND Mathematics Study Panel, 2003); and (c) large proportions of students in this country struggle with algebra. In the meantime, however, findings suggest that an explicit focus on word-problem instruction, designed to support algebraic thinking, is warranted – not only due to the importance of word problems as a predictor of long-term school and employment outcomes (Bynner, 1997; Every Child a Chance Trust, 2009; Parsons & Bynner, 1997; Rivera-Batiz, 1992) but also due to a link between such word-problem instruction and pre-algebraic thinking, as documented in the present study.

Efficacy of Two-Tier Intervention Programs

A secondary purpose of the present study was to examine the efficacy of the calculation and word-problem interventions. Previous studies had validated the *Math Wise* calculation tutoring program and the *Pirate Math* word-problem tutoring program for third-grade students at-risk for poor outcomes (e.g., Fuchs, Powell, et al., 2009; Fuchs, Powell, et al., 2011; Fuchs, Seethaler, et al., 2008). The present study expanded each tutoring program to include a whole-class component and address second-grade skills; ensured alignment between the whole-class and tutoring components; and examined effects of the two-tier RTI systems (whole-class plus tutoring for at-risk students) with second-grade children who spanned the achievement continuum. Results supported the efficacy of both two-tier programs.

Two-tier *Math Wise* produced superior calculation outcomes compared to the business-as-usual control condition and compared to an active contrast condition (*Pirate Math*). This was the case for the proximal calculation outcomes (ES = 0.55 as compared to the control condition and 0.41 as compared to the active contrast condition). It was also the case for the distal calculation outcomes (ES = 0.38 as compared to the control condition and 0.30 as compared to the active contrast condition).

In an analogous way, two-tier *Pirate Math* produced superior word-problem outcomes compared to the business-as-usual control condition and compared to an active contrast condition (*Math Wise*). This was the case for the proximal word problems outcomes (ES = 0.92 as compared to the control condition and 1.00 as compared to the active contrast condition). It was also the case for the proximal and distal pre-algebraic knowledge outcomes (ESs = 1.36 and 0.22 as compared to the control condition and 1.32 and 0.21 as compared to the active contrast condition). Even so, there were no significant effects among the three conditions on the distal word-problem outcome. This finding, which echoes prior work (Fuchs et al., 2009; Fuchs, Powell, et al., 2011), is surprising given that the proximal measure, where significant effects were found, largely mirrors large-scale word-problem tests like the IOWA and was administered in the same fashion as the IOWA. The major differences between the proximal and distal measures were that the proximal measure (a) has more thorough behavior sampling of second-grade word-problem types (IOWA spans grades 1–3; KeyMath spans K-12) and (b) compared to IOWA, requires constructed rather than multiple-choice responses. The most parsimonious explanation for the different pattern of effects on the proximal versus distal word-problem outcome is that students had greater opportunity on the proximal measure to demonstrate the knowledge they had learned during intervention – second-grade word-problem types, with constructed responses.

Study Limitation and Major Conclusions

Before closing, it is important to remind readers that although designed as a cluster randomized design, the present study was carried out as a quasi-experiment primarily because some classroom teachers had previously been randomly assigned to conditions as part of a pilot study; so, they remained in those conditions. Also, 11 of 96 teachers implemented a condition that differed from their assigned condition. As already noted, however, the seriousness of this limitation is mitigated by the fact that, in no case, was condition determined according to teacher, school, or researcher preference. It is also mitigated by supplementary analyses we conducted, which showed that the pattern of results for the subset teachers and classrooms whose assignment was according to design was similar to those reported above, such that the way in which assignment occurred did not alter the effects of conditions. We nevertheless encourage readers to keep this limitation in mind while considering the study’s two major conclusions. First, calculation and word-problem performance at second grade appear to represent distinct aspects of mathematical cognition and indicate the need to address calculation and word-problem performance deliberately and explicitly in research and practice. Second, word-problem intervention, when designed to provide linkages between arithmetic and pre-algebraic thinking, may provide a superior bridge to the pre-algebraic stage of development, compared to arithmetic calculation intervention that is designed to provide linkages.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by Award Number R01 HD059179 and Core Grant #HD15052 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development to Vanderbilt University.

References

Alibali MW, Knuth EJ, Hattikudur S, McNeil NM, Stephens AC. A longitudinal examination of middle school students' understanding of the equal sign and equivalent equations. Mathematical Thinking and Learning. 2007; 9:221–246.10.1080/10986060701360902

Balacheff, N. Symbolic arithmetic vs. algebra: The core of a didactical dilemma. Postscript. In: Sutherland, R.; Rojano, T.; Bell, A.; Lins, RC., editors. Perspectives on school algebra. Dordrecht: The Netherlands: Kluwer; 2001. p. 249-260.

Baroody AJ, Ginsburg HP. The effects of instruction on children's understanding of the "equals" sign. Elementary School Journal. 1983; 84:199–212.10.1086/461356

Blanton, M.; Kaput, J. In: Chick, H.; Stacey, K.; Vincent, J.; Vincent, J., editors. Algebraifying the elementary mathematics experience. Part II: Transforming practice on a district wide scale; The future of the teaching and learning of algebra Proceedings of the 12th ICMI Study Conference; Australia: University of Melbourne; 2001. p. 87-95.

Blanton ML, Kaput JJ. Characterizing a classroom practice that promotes algebraic reasoning. Journal for Research in Mathematics Education. 2005; 36:412–446.

Bynner J. Basic skills in adolescents' occupational preparation. Career Development Quarterly. 1997; 45:305–321.

Carraher, DW.; Schliemann, AD. Is everyday mathematics truly relevant to mathematics education?. In: Moshkovich, J.; Brenner, M., editors. Everyday mathematics Monographs of the Journal for Research in Mathematics Education. Vol. 11. 2002. p. 131-153.

Carpenter, TP.; Levi, L. Developing conceptions on algebraic reasoning in the primary grades. Madison, WI: University of Wisconsin-Madison; 2000. (Report No. 00-2)

Connolly, AJ. KeyMath-Revised. Circle Pines, MN: American Guidance; 1998.

Cummins DD, Kintsch W, Reusser K, Weimer R. The role of understanding in solving word problems. Cognitive Psychology. 1988; 20:405–438.

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

Falkner KP, Levi L, Carpenter TP. Children's understanding of equality: A foundation for algebra. Teaching Children Mathematics. 1999; 6:232–236.

Fuchs LS, Compton DL, Fuchs D, Hollenbeck KN, Craddock C, Hamlett CL. Dynamic assessment of algebraic learning in predicting third graders' development of mathematical problem solving. Journal of Educational Psychology. 2008; 100:829–850. [PubMed: 19884957]

Fuchs LS, Compton DL, Fuchs D, Powell SR, Schumacher RF, Hamlett CL, Vukovic RK. Contributions of domain-general cognitive resources and different forms of arithmetic development to pre-algebraic knowledge. Developmental Psychology. 2012; 48:1315–1326.10.1037/a0027475 [PubMed: 22409764]

Fuchs LS, Fuchs D, Compton DL, Powell SR, Seethaler PM, Capizzi AM, Fletcher JM. The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. Journal of Educational Psychology. 2006; 98:29–43.10.1037/0022-0663.98.1.29

Fuchs LS, Fuchs D, Craddock C, Hollenbeck KN, Hamlett CL, Schatschneider C. Effects of small-group tutoring with and without validated classroom instruction on at-risk students' math problem solving: Are two tiers of prevention better than one? Journal of Educational Psychology. 2008; 100:491–509. [PubMed: 19122881]

Fuchs LS, Fuchs D, Hamlett CL, Phillips NB, Karns K, Dutka S. Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. Elementary School Journal. 1997; 97:223–250.

Fuchs LS, Fuchs D, Stuebing K, Fletcher JM, Hamlett CL, Lambert WE. Problem-solving and computation skills: Are they shared or distinct aspects of mathematical cognition? Journal of Educational Psychology. 2008; 100:30–47.10.1037/0022-0663.100.1.30 [PubMed: 20057912]

Fuchs LS, Geary DC, Compton DL, Fuchs D, Schatschneider C, Hamlett CL, Changas P. Effects of first-grade number knowledge tutoring with contrasting forms of practice. Journal of Educational Psychology. 2013; 105:58–77. [PubMed: 24065865]

Fuchs LS, Geary DC, Compton DL, Fuchs D, Hamlett CL, Bryant JV. The contributions of numerosity and domain-general abilities to school readiness. Child Development. 2010; 81:1520–1533.10.1111/j.1467-8624.2010.01489.x [PubMed: 20840238]

Fuchs LS, Geary DC, Compton DL, Fuchs D, Hamlett CL, Seethaler PM, Schatschneider C. Do different types of school mathematics development depend on different constellations of numerical and general cognitive abilities. Developmental Psychology. 2010; 46:1731–1746.10.1037/a0020662 [PubMed: 20822213]

Fuchs, LS.; Hamlett, CL.; Powell, SR. Second-Grade Calculations Battery. Fuchs, LS., editor. 228 Peabody, Vanderbilt University; Nashville, TN 37203: 2003.

Fuchs LS, Powell SR, Seethaler PM, Cirino PT, Fletcher JM, Fuchs D, Hamlett CL. The effects of strategic counting instruction, with and without deliberate practice, on number combination skill among students with mathematics difficulties. Learning and Individual Differences. 2010; 20:89–100. [PubMed: 20383313]

Fuchs, LS.; Powell, SR.; Seethaler, PM.; Cirino, PT.; Fletcher, JM.; Fuchs, D.; Hamlett, CL. The development of arithmetic and word-problem skill among students with mathematics disability. In: Kauffman, JM.; Hallahan, DP., editors. Handbook of Special Education. New York, NY: Taylor & Francis; 2011. p. 434-444.

Fuchs LS, Powell SR, Seethaler PM, Cirino PT, Fletcher JM, Fuchs D, Zumeta RO. Remediating number combination and word problem deficits among students with mathematics difficulties: A randomized control trial. Journal of Educational Psychology. 2009; 101:561–576. [PubMed: 19865600]

Fuchs, LS.; Seethaler, PM.; Powell, SR. Test of Emerging Algebraic Cognition. Fuchs, LS., editor. 228 Peabody, Vanderbilt University; Nashville, TN 37203: 2009.

Fuchs LS, Seethaler PM, Powell SR, Fuchs D, Hamlett CL, Fletcher JM. Effects of preventative tutoring on the mathematical problem solving of third-grade students with math and reading difficulties. Exceptional Children. 2008; 74:155–173. [PubMed: 20209074]

Fuchs LS, Zumeta RO, Schumacher RF, Powell SR, Seethaler PM, Hamlett CL, Fuchs D. 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. Elementary School Journal. 2010; 110:440–463.10.1086/651191

Fuson KC, Kwon Y. Korean children's single-digit addition and subtraction: Numbers structured by ten. Journal for Research in Mathematics Education. 1992; 23:148–165.

Geary, DC.; Boykin, AW.; Embretson, S.; Reyna, V.; Siegler, R.; Berch, DB.; Graban, J. The Final Report of the National Mathematics Advisory Panel. Washington, DC: U.S. Department of Education; 2008. Report of the task group on learning processes.

Greenes, C.; Larson, M.; Leiva, MA.; Shaw, JM.; Stiff, L.; Vogeli, BR.; Yeatts, K. Boston, MA: Houghton Mifflin; 2005.

Groen G, Resnick LB. Can preschool children invent addition algorithms? Journal of Educational Psychology. 1977; 69:645–652.

Herscovics N, Kieran C. Construction meaning for the concept of equation. Mathematics Teacher. 1980:572–580.

Jacobs VR, Franke ML, Carpenter TP, Levi L, Battey D. Professional development focused on children's algebraic reasoning in elementary school. Journal for Research in Mathematics Education. 2007; 38:258–288.

Hox, JJ. Multilevel analysis: Techniques and applications. 2. New York: Routledge; 2010.

Hoover, HD.; Hieronymous, AN.; Dunbar, SB.; Frisbie, DA. Iowa Test of Basic Skills, Form K. Itasca, IL: Riverside; 1993.

Jacobs VR, Franke ML, Carpenter TP, Levi L, Battey D. Professional development focused on children's algebraic reasoning in elementary school. Journal for Research in Mathematics Education. 2007; 38:258–288.

Jitendra AK, Star JR, Rodriguez M, Lindell M, Someki F. Improving students' proportional thinking using schema-based instruction. Learning and Instruction. 2011; 21:731–745.

Jitendra AK, Star JR, Starosta K, Leh JM, Sood S, Caskie G, Hughes CL, Mack TR. Improving seventh grade students' learning of ratio and proportion: The role of schema-based instruction. Contemporary Educational Psychology. 2009; 34:250–264.

Jordan NC, Hanich L. Mathematical thinking in second-grade children with different forms of LD. Journal of Learning Disabilities. 2000; 33:567–578.10.1177/002221940003300605 [PubMed: 15495398]

Kieran, C. Cognitive processes involved in learning school algebra. In: Nesher, P.; Kilpatrick, J., editors. Mathematics and cognition: a research synthesis by the International Group for the Psychology of Mathematics Education. New York: Cambridge University Press; 1990. p. 96-112.

Kieran, C. The learning and teaching of school algebra. In: Grouws, DA., editor. Handbook of research on mathematics teaching and learning. Reston, VA: National Council of Teachers of Mathematics; 1992. p. 390-419.

Kintsch W, Greeno JG. Understanding and solving word arithmetic problems. Psychological Review. 1985; 92:109–129. [PubMed: 3983303]

Knuth EJ, Stephens AC, McNeil NM, Alibali MW. Does understanding the equal sign matter? Evidence from solving equations. Journal for Research in Mathematics Education. 2006; 37:297–312.

Lee K, Ng SF, Bull R, Pe ML, Ho RHM. Are patterns important? An investigation of the relationships between proficiency in patterns, computation, executive functioning, and algebraic word problems. Journal of Educational Psychology. 2011; 103:269–281.10.1037/a0023068

LeFevre JA, Morris J. More on the relation between division and multiplication in simple arithmetic: Evidence for mediation of division solutions via multiplication. Memory & Cognition. 1999; 27:803–812. [PubMed: 10540809]

Linchevski, L. In: Heuvel-Panhuizen, Mvd, editor. Operating on the unknowns: What does it really mean?; Proceedings of the 25th Conference of the International Group for the Psychology of Mathematics Education; 2001. p. 141-144.

Linchevski L, Herscovics N. Crossing the cognitive gap between arithmetic and algebra: Operating on the unknown in the context of equations. Educational Studies in Mathematics. 1996; 30:39–65.10.1007/BF00163752

Littell, RC.; Milliken, GA.; Stroup, WW.; Wolfinger, RD.; Schabenberger, O. SAS for mixed models. 2. Cary, NC: SAS Institute; 2006.

Luke, DA. Multilevel modeling (quantitative applications in the social sciences). Thousand Oaks, CA: Sage; 2004.

Ma, X.; Ma, L. Using multilevel modeling to investigate school effects. In: O'Connell, AA.; McCoach, DB., editors. Multilevel modeling of educational data. Charlotte, NC: Information Age Publishing; 2008. p. 59-110.

McCoach, DB.; Black, AC. Evaluation of model fit and adequacy. In: O'Connell, AA.; McCoach, DB., editors. Multilevel modeling of educational data. Charlotte, NC: Information Age Publishing; 2008. p. 245-272.

McGregor K, Newman R, Reilly R, Capone N. Semantic representation and naming in children with specific language impairment. Journal of Speech, Language, and Hearing Research. 2002; 45:998–1014.

McNeil NM. Limitations to teaching children 2 + 2 = 4: Typical arithmetic problems can hinder learning of mathematical equivalence. Child Development. 2008; 79:1524–1537. [PubMed: 18826540]

McNeil NM, Alibali MW. Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. Child Development. 2005; 76:883–899.10.1111/j.1467-8624.2005.00884.x [PubMed: 16026503]

Nathan MJ, Kintsch W, Young E. A theory of algebra-word-problem comprehension and its implications for the design of learning environments. Cognitive and Instruction. 1992; 9:392–389.

National Mathematics Advisory Panel. Final report. Washington, DC: U.S. Department of Education; 2008.

Parsons S, Bynner J. Numeracy and employment. Education and Training. 1997; 39:43–51.

Peugh JL. A practical guide to multilevel modeling. Journal of School Psychology. 2010; 48:85–112. [PubMed: 20006989]

Pillay H, Wilss L, Boulton-Lewis G. Sequential development of algebra knowledge: A cognitive analysis. Mathematics Education Research Journal. 1998; 10(2):87–102.10.1007/BF03217344

Powell SR. Equations and the equal sign in elementary mathematics textbooks. Elementary School Journal. 2012; 112:627–648. [PubMed: 22707760]

Powell SR, Fuchs LS. Contribution of equal-sign instruction beyond word-problem tutoring for third-grade students with mathematics difficulty. Journal of Educational Psychology. 2010; 102:381–394.10.1037/a0018447 [PubMed: 20640240]

RAND Mathematics Study Panel. Mathematical proficiency for all students: Toward a strategic research and development program in mathematics education (No. 08330331X). Santa Monica, CA: RAND; 2003.

Raudenbush, SW.; Bryk, AS. Hierarchical linear models. Thousand Oaks, CA: Sage; 2002.

Rittle-Johnson B, Alibali MW. Conceptual and procedural knowledge of mathematics: Does one lead to the other? Journal of Educational Psychology. 1999; 91:175–189.10.1037/0022-0663.91.1.175

Rivera-Batiz FL. Quantitative literacy and the likelihood of employment among young adults in the United States. The Journal of Human Resources. 1992; 27:313–328.

SAS Institute. SAS 9.3. Cary, NC: Author; 2002–201.

Sfard A. On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. Educational Studies in Mathematics. 1991; 22:1–16.10.1007/BF00302715

Sfard A, Linchevski L. The gains and the pitfalls of reification – the case of algebra. Educational Studies in Mathematics. 1994; 26:191–228.10.1007/BF01273663

Siegler, RS.; Shrager, J. Strategy choice in addition and subtraction: How do children know what to do?. In: Sophian, C., editor. Origins of cognitive skills. Hillsdale, NJ: Erlbaum; 1984. p. 229-293.

Snijders, TAB.; Bosker, RJ. Multilevel analysis: An introduction to basic and advanced multilevel modeling. 2. London: Sage; 2012.

Sullivan LM, Dukes KA, Losina E. Tutorial in biostatistics: An introduction to hierarchical linear modeling. Statistics in Medicine. 1999; 18:855–888. [PubMed: 10327531]

Swanson HL. Cross-sectional and incremental changes in working memory and mathematical problem solving. Journal of Educational Psychology. 2006; 98:265–281.10.1037/0022-0663.98.2.265

Swanson HL, Jerman O, Zheng X. Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. Journal of Educational Psychology. 2008; 100:343–379.

Tolar TD, Lederberg AR, Fletcher JM. A structural model of algebra achievement: Computational fluency and spatial visualization as mediators of the effect of working memory on algebra achievement. Educational Psychology. 2009; 29:239–266.10.1080/01443410802708903

Tymms, P. Effect sizes in multilevel models. In: Schagen, I.; Elliot, K., editors. But what does it mean?: The use of effect sizes in educational research. Slough, Berks: National Foundation for Educational Research; 2004. p. 55-66.

U.S. Department of Education. Digest of Education Statistics. 2008. Retrieved from http://nces.ed.gov/pubs2009/2009020.pdf

Wechsler, D. Wechsler Abbreviated Scale of Intelligence. San Antonio, TX: Psychological Corporation; 1999.

Wilkinson, GS. Wide Range Achievement Test 3. Wilmington, DE: Wide Range; 1993.

Woodcock, RW. Woodcock Diagnostic Reading Battery. Itasca, IL: Riverside; 1997.

Zhu, J. WASI manual. San Antonio: The Psychological Corporation; 1999.

Table 1

Performance Data by Treatment Condition

	Condition											
	CAL				WP				Control			
	Pre		Post		Pre		Post		Pre		Post	
Variable	**X**	**(SD)**	**X**	**(SD)**	**X**	**(SD)**	**X**	**(SD)**	**X**	**(SD)**	**X**	**(SD)**
Screening												
Sums to 12	8.26	(4.59)	NA		8.27	(4.59)	NA		8.56	(4.88)	NA	
Story Problems	6.99	(3.79)	NA		7.09	(3.61)	NA		6.64	(3.45)	NA	
WASI IQ	93.42	(12.98)	NA		94.96	(13.66)	NA		93.30	(12.19)	NA	
Calculation Outcomes												
Proximal												
Sums to 18	6.23	(3.54)	12.29	(4.89)	6.03	(3.70)	10.66	(4.86)	5.86	(3.77)	9.92	(5.08)
Minuends to 12	4.37	(2.56)	8.74	(5.21)	4.37	(2.77)	8.09	(4.20)	4.30	(2.71)	6.81	(4.27)
Minuends to 18	3.19	(2.53)	7.61	(4.46)	3.00	(2.59)	6.54	(4.23)	2.92	(2.39)	5.54	(4.27)
2-Digit Addition	3.53	(3.60)	15.46	(4.93)	3.15	(3.63)	11.28	(5.76)	3.43	(3.26)	11.55	(6.40)
2-Digit Subtraction	2.06	(2.42)	8.42	(5.02)	1.96	(2.43)	2.37	(2.71)	5.60	(5.34)	6.39	(4.20)
Distal												
WRAT-Raw -Standard	19.36	(2.62)	23.64	(1.98)	19.35	(2.53)	22.46	(2.48)	19.15	(2.60)	22.46	(2.48)
	94.31	(12.34)	102.87	(9.79)	94.53	(12.39)	99.97	(10.67)	93.74	(13.05)	98.32	(11.83)
Key Math Addition-Raw-Standard	7.32	(1.78)	9.62	(2.11)	6.85	(1.92)	8.82	(2.21)	6.84	(1.96)	8.60	(2.34)
	104.35	(11.52)	109.01	(12.74)	100.90	(11.10)	104.83	(12.81)	101.23	(11.73)	103.75	(13.61)
Key Math Subtraction-Raw-Standard	4.32	(1.87)	6.01	(2.21)	4.41	(2.16)	5.51	(2.27)	4.12	(1.96)	5.39	(2.35)
	100.74	(13.02)	103.25	(13.71)	100.21	(14.19)	100.73	(13.86)	99.03	(13.79)	99.95	(14.82)
Word-Problem Outcomes												
Proximal												
2nd Grade Story Problems	8.13	(4.88)	12.05	(6.00)	8.31	(5.19)	18.44	(7.82)	7.41	(4.34)	12.28	(6.11)
Distal												
Key Math Prob Solving-Raw-Standard	2.42	(1.62)	3.61	(1.96)	2.49	(2.00)	3.94	(2.08)	2.37	(1.71)	3.62	(1.90)
	104.16	(9.82)	105.74	(10.95)	104.44	(9.66)	107.59	(11.12)	103.71	(10.49)	105.66	(10.53)
Iowa-Raw-Standard	NA		16.77	(5.07)	NA		16.73	(5.09)	NA		16.48	(4.65)

	Condition											
	CAL				WP				Control			
	Pre		Post		Pre		Post		Pre		Post	
Variable	**X**	**(SD)**	**X**	**(SD)**	**X**	**(SD)**	**X**	**(SD)**	**X**	**(SD)**	**X**	**(SD)**
	NA		157.59	(17.48)	NA		157.50	(17.69)	NA		156.47	(15.61)
Pre-Algebra												
Proximal												
Find X	NA		4.32	(2.83)	NA		6.80	(1.69)	NA		4.42	(2.96)
Number Sentences	NA		1.40	(1.69)	NA		3.57	(2.60)	NA		1.18	(1.69)
Distal												
Dynamic Assessment	NA		7.92	(4.10)	NA		8.67	(4.47)	NA		7.93	(4.49)
Test of Pre-Algebra	NA		10.96	(5.41)	NA		12.01	(5.18)	NA		10.61	(5.24)

Table 2

Model Characteristics for Proximal and Distal Calculation Outcomes

	Proximal					Distal				
Parameter	**Estimate**	**(S.E.)**	***Df***	***t/z/F***	**95% CI**	**Parameter Estimate**	**(S.E.)**	***df***	***t/z/F***	**95% CI**
Unconditional Model										
Fixed: Intercept (γ_{000})	−0.03	(0.07)	22.4	−0.48	−0.17 to 0.11	−0.02	(0.06)	22.3	−0.29	−0.13 to 0.10
Random: Teacher(σ^2_{u0})	0.10	(0.03)		3.85[c]	0.07 to 0.18	0.05	(0.02)		2.76[b]	0.03 to 0.12
School (σ^2_{v0})	0.07	(0.03)		1.97[a]	0.03 to 0.25	0.05	(0.02)		2.04[a]	0.02 to 0.16
Residual (σ^2_{e})	0.71	(0.03)		22.23[c]	0.65 to 0.78	0.63	(0.03)		22.26[c]	0.58 to 0.69
Full/Final Model										
Fixed: Intercept (γ_{000})	−0.28	(0.11)	38.2	−2.48[a]	−0.50 to − 0.05	0.08	(0.11)	74.5	0.73	−0.14 to 0.31
Calculations Pretest	0.54	(0.03)	1010	20.14[c]	0.49 to 0.59	0.35	(0.03)	1013	12.85[c]	0.30 to 0.40
Tutoring Eligibility	0.07	(0.06)	1010	1.16	−0.05 to 0.18	−0.17	(0.06)	1007	−2.90[b]	−0.29 to − 0.06
Sex (Female)			NA			0.09	(0.04)	970	2.12[a]	0.01 to 0.18
Lunch (Free/Reduced)			NA			−0.23	(0.07)	915	−3.39[b]	−0.37 to − 0.10
Ethnicity (A Amer)			NA			−0.21	(0.06)	922	−3.41[c]	−0.33 to − 0.09
Ethnicity (Hispanic)			NA			0.04	(0.07)	923	0.54	−0.10 to 0.18
Ethnicity (Other)			NA			0.11	(0.10)	990	1.11	−0.09 to 0.31
Treatment (CAL)	0.42	(0.09)	98.7	4.71[c]	0.25 to 0.60	0.36	(0.09)	101	4.27[c]	0.20 to 0.53
Treatment (WP)	0.11	(0.09)	95.6	1.23	−0.07 to 0.28	0.14	(0.08)	98.1	1.59	−0.03 to 0.30
Cohort (1)	−0.10	(0.13)	29.2	−0.76	−0.36 to 0.16	0.01	(0.11)	27.6	0.13	−0.20 to 0.23
Cohort (2)	0.06	(0.13)	29.9	0.45	−0.20 to 0.32	−0.01	(0.10)	28.2	−0.08	−0.22 to 0.20
Cohort (3)	0.25	(0.13)	30.6	1.89	−0.02 to 0.51	0.03	(0.11)	27.2	0.29	−0.19 to 0.25
Random: Teacher (σ^2_{u0})	0.08	(0.03)		2.41[b]	0.04 to 0.23	0.05	(0.02)		2.93[b]	0.03 to 0.10
School (σ^2_{v0})	0.04	(0.02)		1.81[a]	0.02 to 0.17	0.02	(0.01)		1.17	0.01 to 0.28

	Proximal					Distal				
Parameter	**Estimate**	**(S.E.)**	***Df***	***t/z/F***	**95% CI**	**Parameter Estimate**	**(S.E.)**	***df***	***t/z/F***	**95% CI**
Residual (σ_e^2)	0.47	(0.02)		21.52[c]	0.43 to 0.52	0.48	(0.02)		21.43[c]	0.44 to 0.53

Note.

[a] $p < .05$;

[b] $p < .01$;

[c] $p < .001$.

CAL is calculation intervention. WP is word-problem intervention. Proximal and distal outcomes are standardized composites. Covariance terms are not shown. For the unconditional model, *t* values are reported. For full/final model, *F*-values are reported for the fixed portion of model; *z*-values for the random portion of model. For *df*s, numerators (not included) are all 1; denominators (reported) use the Kenward-Roger approximation. *Under Fixed Effects*, intercept refers to the outcome variable. Pretest is grand-mean centered. The code 0 is for the treatment control group, for no free lunch, for Caucasian, for ineligible for tutoring, and for cohort 4. For meaning of symbols under *Random Effects*, see text – although symbols in the table stand for variance components (e.g., σ_e^2 denotes the student-level residual variance).

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 3

Model Characteristics for Proximal and Distal Word-Problem Outcomes

	Proximal					Distal				
Parameter	**Estimate**	**(S.E.)**	***df***	***t/ z/F***	**95% CI**	**Parameter Estimate**	**(S.E.)**	***df***	***t/ z/F***	**95% CI**
Unconditional Model										
Fixed: Intercept (γ_{000})	14.44	(0.45)	21.6	32.28[c]	13.51 to 15.37	−0.03	(0.06)	23.5	−0.49	−0.15 to 0.09
Random: Teacher (σ^2_{u0})	14.35	(2.94)		4.88[c]	9.96 to 22.44	0.00	(0.01)		0.27	0.00 to 1.09
School (σ^2_{v0})	3.62	(3.05)		0.33	0.05 to 2.69	0.06	(0.02)		2.71[b]	0.03 to 0.15
Residual (σ^2_{e})	41.47	(1.87)		22.23[c]	38.04 to 45.38	0.58	(0.03)		22.19[c]	0.53 to 0.63
Full/Final Model										
Fixed: Intercept (γ_{000})	11.38	(0.75)	77.6	15.11[c]	9.89 to 12.88	0.17	(0.07)	81	2.23[b]	0.02 to 0.33
Word-Problem Pretest	5.09	(0.26)	1028	19.81[c]	4.59 to 5.60	0.73	(0.03)	1053	28.75[c]	0.68 to 0.78
Tutoring Eligibility	−1.01	(0.43)	1061	−2.32[a]	−1.86 to −0.16	−0.20	(0.04)	1044	−4.94[c]	−0.28 to −0.12
Sex (Female)	1.00	(0.33)	1039	3.05[b]	0.36 to 1.64	−0.07	(0.03)	1043	−2.32[a]	−0.13 to −0.01
Lunch (Free/Reduced)			NA			−0.12	(0.05)	957	−2.50[a]	−0.21 to −0.03
Ethnicity (A Amer)			NA			−0.19	(0.04)	970	−4.35[c]	−0.28 to −0.11
Ethnicity (Hispanic)			NA			0.04	(0.06)	1014	0.66	−0.07 to 0.15
Ethnicity (Other)			NA			0.02	(0.07)	1040	0.28	−0.12 to 0.16
Second Language			NA			−0.20	(0.06)	1055	−3.29[c]	−0.32 to −0.08
Treatment (CAL)	0.78	(1.00)	145	0.78	−1.19 to 2.75	0.05	(0.05)	1046	1.18	−0.13 to 0.17
Treatment (WP)	8.99	(1.02)	149	8.83[c]	6.98 to 11.00	0.08	(0.04)	1055	1.76	−0.01 to 0.17
Cohort (1)	0.80	(1.07)	88.9	0.75	−1.33 to 2.93	0.12	(0.07)	33	1.69	0.02 to 0.27
Cohort (2)	1.22	(1.03)	90.1	1.19	−0.83 to 3.27	0.11	(0.07)	33.8	1.57	0.03 to 0.26
Cohort (3)	0.64	(1.12)	108	0.58	−1.57 to 2.86	0.01	(0.07)	34.7	0.19	−0.14 to 0.17
Random: Teacher (σ^2_{u0})	0.97	(0.59)		1.64	0.39 to 5.33	0	(--)		--	--
School (σ^2_{v0})	0.37	(0.48)		0.77	0.08 to 134.18	0.01	(0.01)		2.19[a]	0.01 to 0.04

	Proximal					Distal				
Parameter	**Estimate**	**(S.E.)**	***df***	***t/ z/F***	**95% CI**	**Parameter Estimate**	**(S.E.)**	***df***	***t/ z/F***	**95% CI**
Residual (σ_e^2)	27.70	(1.26)		22.04[c]	25.39 to 30.34	0.26	(0.01)		22.74[c]	0.24 to 0.28

See note for Table 2. Parameter estimates for the significant treatment by cohort interaction for Proximal Word Problems are not shown in the table, but were included in the model from which the above numbers were derived, and are described in text.

Table 4

Model Characteristics for Proximal and Distal Pre-Algebra Outcomes

	Proximal					Distal				
Parameter	**Estimate**	**(S.E.)**	***df***	***t/z/F***	**95% CI**	**Parameter Estimate**	**(S.E.)**	***df***	***t/z/F***	**95%**
Unconditional Model										
Fixed: Intercept (γ_{000})	−0.01	(0.04)	108	−0.14	−0.09 to 0.07	−0.04	(0.05)	22.9	−0.82	−0.14 to 0.06
Random: Teacher(σ^2_{u0})	0.15	(0.02)		6.01[c]	0.11 to 0.21	0.01	(0.01)		0.98	0.00 to 0.59
School (σ^2_{v0})	0	--				0.05	(0.02)		2.07[a]	0.02 to 0.13
Residual (σ^2_{e})	0.28	(0.01)		22.22[c]	0.25 to 0.30	0.78	(0.04)		22.41[c]	0.72 to 0.86
Full/Final Model										
Fixed: Intercept (γ_{000})	−0.18	(0.08)	122	−2.33 [a]	−0.33 to −0.03	0.02	(0.08)	96.3	0.20	−0.15 to 0.19
Word-Problem Pretest	0.27	(0.03)	1006	10.30[c]	0.22 to 0.32	0.60	(0.04)	1012	15.72[c]	0.52 to 0.67
Calculations Pretest	0.06	(0.02)	1004	3.27[c]	0.03 to 0.10	0.21	(0.03)	1014	7.56[c]	0.16 to 0.27
Tutoring Eligibility	−0.04	(0.04)	1007	−0.99	−0.12 to 0.04	−0.13	(0.06)	1009	−2.16[a]	−0.24 to −0.01
Sex (Female)						−0.10	(0.04)	990	−2.31[a]	−0.18 to −0.01
Lunch (Free/Reduced)	−0.09	(0.04)	758	−2.02[a]	−0.18 to −0.00					
Second Language	−0.14	(0.06)	897	−2.48[a]	−0.26 to −0.03	−0.18	(0.08)	905	−2.18[a]	−0.34 to −0.02
Ethnicity (A Amer)	−0.08	(0.04)	820	−1.82	−0.16 to 0.01	−0.15	(0.06)	808	−2.64[b]	−0.27 to −0.04
Ethnicity (Hispanic)	0.11	(0.05)	846	2.01[a]	0.00 to 0.21	0.19	(0.08)	840	2.55[a]	0.04 to 0.34
Ethnicity (Other)	0.10	(0.07)	933	1.46	−0.03 to 0.23	0.16	(0.10)	922	1.62	−0.03 to 0.35
Treatment (CAL)	−0.00	(0.09)	117	−0.05	−0.19 to 0.18	0.03	(0.07)	115	0.43	−0.11 to 0.17
Treatment (WP)	0.74	(0.09)	120	7.81[c]	0.55 to 0.92	0.16	(0.07)	116	2.29[a]	0.02 to 0.30
Cohort (1)	0.15	(0.10)	93.9	1.52	−0.05 to 0.34	−0.13	(0.08)	41.4	−1.55	−0.29 to 0.04
Cohort (2)	0.01	(0.09)	98.6	0.16	−0.17 to 0.20	0.14	(0.08)	43.5	1.70	−0.03 to 0.30
Cohort (3)	−0.14	(0.10)	123	−1.36	−0.34 to 0.06	0.04	(0.08)	41.1	0.53	−0.13 to 0.22
Random: Teacher (σ^2_{u0})	0.01	(0.01)		1.82	0.00 to 0.05	0.02	(0.01)		1.86[a]	0.01 to 0.08
School (σ^2_{v0})	0.00	(0.00)		0.56	0.00 to 125.06	0.01	(0.01)		0.92	0.00 to 0.42

	Proximal					Distal				
Parameter	**Estimate**	**(S.E.)**	***df***	***t/z/F***	**95% CI**	**Parameter Estimate**	**(S.E.)**	***df***	***t/z/F***	**95%**
Residual (σ_e^2)	0.21	(0.01)		21.32[c]	0.20 to 0.23	0.45	(0.02)		21.58[c]	0.41 to 0.49

See note for Table 2. Parameter estimates for the significant treatment by cohort interaction for Proximal Pre-Algebra are not shown in the table, but were included in the model from which the above numbers were derived, and are described in text.