# Effects of a Multitier Support System on Calculation, Word Problem, and Prealgebraic Performance Among At-Risk Learners

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

[1]University of Texas at Austin

[2]Vanderbilt University

[3]University of Houston

[4]Metropolitan Nashville Public Schools

## Abstract

The focus of the present study was enhancing word-problem and calculation achievement in ways that support pre-algebraic thinking among 2nd-grade students at risk for mathematics difficulty. Intervention relied on a multi-tier support system (i.e., responsiveness-to-intervention or RTI) in which at-risk students participate in general classroom instruction and receive supplementary small-group tutoring. Participants were 265 students in 110 classrooms in 25 schools. Teachers were randomly assigned to 3 conditions: calculation RTI, word-problem RTI, and business-as-usual control. Intervention lasted 17 weeks. Multilevel modeling indicated that calculation RTI improved calculation but not word-problem outcomes; word-problem RTI enhanced proximal word-problem outcomes as well as performance on some calculation outcomes; and word-problem RTI provided a stronger route than calculation RTI to pre-algebraic knowledge.

## Keywords

mathematics; calculation; word problems; pre-algebra; Response to Intervention

---

Calculations and word problems are critical forms of mathematics competence for success in the primary grades and through adulthood. A calculation (CAL) problem is set up for solution. By contrast, a word problem (WP) requires students to process text to build a problem model and then construct a number sentence for calculating the unknown. Along with this transparent difference in the nature of these tasks, correlational studies suggest that

different cognitive abilities underlie CAL and WP skill (e.g., Fuchs, Fuchs, Stuebing, et al., 2008; Fuchs, Geary, et al., 2010; Swanson, 2006).

At the same time, students who have concurrent difficulty in CAL and WPs have more severe deficits in each domain (e.g., Fuchs, Fuchs, Stuebing, et al., 2008), and they experience more pervasive challenges on other mathematics competencies. This includes pre-algebraic thinking (Powell & Fuchs, 2014), which is problematic because algebra is frequently required for high-school graduation (U.S. Department of Education, 2008), is associated with successful participation in the U.S. workforce, and represents a gateway to more advanced mathematics achievement (National Mathematics Advisory Panel [NMAP], 2008; RAND Mathematics Study Panel, 2003). Interest has therefore increased in identifying ways to design early mathematics instruction to create a strong platform for algebraic reasoning (see, for example, NMAP, 2008).

For these reasons, our focus in the present study was enhancing second-grade CAL and WP achievement in ways that simultaneously support pre-algebraic thinking. Our target population was students with strong risk for mathematics difficulty (MD): those with concurrent low performance at the start of second grade on CAL and WPs (i.e., CAL/WP-MD). Our approach to intervention relied on a multi-tier support system, often referred to as *responsiveness-to-intervention* (RTI; Vaughn & Fuchs, 2003), in which at-risk students participate in general classroom instruction while receiving supplementary small-group tutoring (Fuchs, Fuchs, Craddock, et al., 2008; Lembke, Garman, Deno, & Stecker, 2010; Mellard, Stern, & Woods, 2011). With such tutoring, schools provide students with a learning environment to remediate foundational skills by explaining mathematical concepts and procedures using more simple and direct language, by providing practice in more systematic and novel ways, and by delivering more immediate corrective feedback when students demonstrate incorrect mathematical thinking. In providing this second layer of support, the hope is that small-group tutoring works synergistically with classroom mathematics instruction to strengthen the learning of students with MD.

Importantly, however, we were also interested in whether a focus on CAL or WPs provides a stronger route toward pre-algebraic thinking for students with CAL/WP-MD. We therefore designed and implemented two separate multi-level mathematics systems, one in CAL and another in WPs and randomly assigned teachers and their classrooms to WP-RTI, CAL-RTI, or business-as-usual control. In this introduction, we provide background information on pre-algebra and its possible connections to CAL and WPs. We then outline prior work related to CAL-RTI or WP-RTI and explain the present study's purpose and hypotheses. Note that a parent research report (Fuchs et al., in press) also focused on transfer from CAL and from WP intervention to pre-algebraic thinking. That report considered a larger sample of students who received one or two levels of support, whose mathematics skills spanned the continuum, with difficulty in calculations and/or word problems, or in neither area. That report did not disaggregate outcomes according to difficulty status.

## Pre-Algebraic Thinking and Possible Connections to Calculations and Word Problems

In contrast to arithmetic, algebra involves symbolizing and operating on numerical relationships within mathematical structures. Some subscribe to the position that algebra creates an interference effect with arithmetic, and algebra is therefore developmentally inappropriate for young students (e.g., Balacheff, 2001; Linchevski, 2001; Linchevski & Herscovics, 1996; Sfard, 1991). Yet, because algebraic expressions can be treated procedurally, by substituting numerical values to yield numerical results (Kieran, 1990), understanding of arithmetic principles appears to involve generalizations that are algebraic in nature. This suggests algebra warrants a role in early instruction (Blanton & Kaput, 2005; Carraher & Schliemann, 2002; NMAP, 2008). A third view, which bridges both perspectives, supports a connection between arithmetic and algebraic thinking, but only if arithmetic instruction – on whole-number calculations or word problems – is designed to facilitate that transition.

Pillay, Wilss, and Boulton-Lewis's (1998) model of learning underpins the third perspective, and Kieran (1990) relied on Pillay et al. to propose a developmental progression. The first stage is arithmetic competence: the capacity to operate numerically and the understanding of operational laws and relational meaning of the equal sign in standard equations (i.e., both sides of the equal sign have the same value). This provides the foundation for a pre-algebraic stage that builds on arithmetic competence by expanding the relational meaning of the equal sign to include nonstandard equations (e.g., $4 = 13 - 9$; $4 + 8 = 6 + 6$), the concept of unknowns in equations, and the concept of a variable. This middle stage supports development of formal algebraic competence and indicates potential targets for early pre-algebraic intervention: understanding of the equal sign (i.e., solving nonstandard equations with one unknown) and the concept of a variable (i.e., completing function tables). It also suggests the need for instruction that supports the transition between arithmetic competence and formal algebraic competence.

Unfortunately, research illustrates how conventional instruction fails to address this need. For example, typical instruction causes most students to misconstrue the equal sign as an operational symbol (Baroody & Ginsburg, 1983; McNeil & Alibali, 2005; Powell, 2012). For example, students commonly answer problems like $7 + 5 = \_\_ + 3$ as 12 (ignoring the operation to the right side of the equal sign) or 15 (adding 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).

It is nevertheless unclear whether a focus on arithmetic CAL or arithmetic WPs is more productive for facilitating the transition from arithmetic to algebraic thinking. Fluency with CAL may reflect a strong foundation in arithmetic operational laws and may 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, WPs require CAL as well as two forms of symbolic representations (numerals

and language), even as word problems reflect understanding of relationships between known and unknown quantities (Geary et al., 2008). Word problems may therefore involve greater symbolic complexity than CAL and 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).

Lee, Ng, Bull, Pe, and Ho (2011) and Tolar, Lederberg, and Fletcher (2009) have investigated potential connections with algebra, but these studies are correlational and investigated either CAL or WPs. We located only one report (Fuchs et al., 2012) that contrasted the connection to algebra of WPs versus CAL. Although results suggested WPs have a stronger connection, the design was again correlational. What is required is an experimental study examining whether intervention in CAL or WPs improves students' pre-algebraic knowledge. In the present study, we focused on students with CAL/WP-MD – a challenging population for expecting arithmetic intervention to result in superior pre-algebraic cognition.

## Prior Work Related to CAL or WP Small-Group Tutoring or RTI

Previous studies have assessed the efficacy of small-group tutoring to improve the mathematics performance of students with MD, typically defined on a measure of CAL or a broad-based mathematics achievement test. An instructional model that incorporates explicit instruction with multiple visual representations, heuristics, and student verbalizations contributes to improved mathematics performance in this population of learners (Gersten et al., 2009). Most but not all tutoring studies that incorporate such practices have been associated with positive effects (e.g., Fuchs, Powell, et al., 2008; Jitendra, Hoff, & Beck, 1999; Powell & Fuchs, 2010; Tournaki, Bae, & Kerekes, 2008). For example, Bryant, Bryant, Gersten, Scammacca, Funk, et al. (2008) conducted a regression-discontinuity study with first graders who were identified for tutoring due to low performance on a mathematics progress-monitoring measure. Tutoring focused on number recognition, counting, comparing numbers, place value, and number combinations. At posttest, students scored significantly higher than expected. By contrast, Bryant, Bryant, Gersten, Scammacca, and Chavez (2008) found mixed results; second graders, but not first graders, showed performance gains as a function of similarly designed tutoring. Moreover, at posttest, the second graders did not perform comparably to students without MD. The authors thus concluded that more powerful interventions are needed to realize the necessary improvement for this population of learners.

One strategy for increasing intervention power may reside in a multi-tier framework, such as RTI (Vaughn & Fuchs, 2003), in which students with CAL/WP-MD receive research-based classroom instruction in combination with well-designed small-group tutoring. Yet, RTI is a relatively new innovation, and insufficient attention has been directed at understanding the potential of mathematics RTI for students with MD. We identified only one randomized control trial focused on a multi-tier support system in mathematics. At third grade, Fuchs, Fuchs, Craddock, et al. (2008) randomly assigned classrooms to receive research-principled WP whole-class instruction or to receive the school's standard WP instruction. Then, within

each classroom, at-risk students were identified due to poor WP performance and randomly assigned to receive supplementary small-group tutoring support or to continue in the general education program without such support. At posttest, students who received the combination of research-principled whole-class and small-group WP instruction demonstrated substantially better learning than students in the other three conditions. This led the authors to conclude that two tiers of support are better than one tier for students with MD.

In the present study, we therefore focused on the efficacy of two-tier mathematics RTI systems that incorporate whole-class research-principled general education instruction, combined with supplementary small-group instruction. We extended the literature in four ways: (a) by target a population not previously studied in prior RTI or tutoring studies, students with CAL/CP-MD; (b) by focusing on second grade, when MD is established but when intervention may be early enough to reverse the trajectory of mathematics development; (c) contrasting the efficacy of a multi-tier system focused on CAL skill against an analogous system focused on WP skill, while also including a business-as-usual control condition; and (d) by examining whether a focus on CAL or WPs within an RTI provides a stronger route toward pre-algebra thinking.

Note that our major purpose was to examine effects of the two-tier intervention systems on calculations, word problems, and pre-algebraic knowledge. Our major contrast, therefore, was between the two active RTI conditions. At the same time, including a BAU 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. (See Instructional Framework in Method section for information on the instructional approach taken in each RTI system and its potential connection with pre-algebra.)

## Study Purpose and Hypotheses

The major purpose of this study was, therefore, to assess the efficacy of two RTI programs, one focused on CAL and the other on WPs, specifically for second-grade students with CAL/WP-MD. We contrasted the efficacy of CAL-RTI and WP-RTI against each other and against the control group on CAL, WPs, and pre-algebraic knowledge. CAL intervention provided no instruction on WPs; WP intervention provided no instruction on CAL (students were told to use whatever methods their classroom teachers had taught to calculate answers). The second-grade two-tiered CAL intervention was based in part on previously validated whole-class (Fuchs et al., 1997) and tutoring (Fuchs et al., 2011; Fuchs, Powell, et al., 2009) programs at third grade. The second-grade two-tiered WP intervention was based in part on previously validated whole-class (Fuchs, Fuchs, Craddock, et al., 2008) and tutoring (Fuchs et al., 2011; Fuchs, Powell, et al., 2009) programs at third grade. These programs are referred to, respectively, as *Math Wise* and *Pirate Math*.

Our hypotheses were as follows. First, we expected effects to be specific to the focus of intervention: that CAL-RTI would produce superior CAL outcomes, compared to WP-RTI and to control, and that WP-RTI would result in superior WP outcomes, compared to CAL-RTI and to control. Second, because we designed intervention to facilitate the transition

from arithmetic to algebra, we expected CAL-RTI and WP-RTI to improve pre-algebraic thinking, with both conditions outperforming the control group. We did, however, expect WP-RTI to produce stronger pre-algebraic performance than CAL-RTI, in light of Fuchs et al.'s (2012) correlational study and based on the assumption that arithmetic WPs rely more on the type of mental flexibility, manipulation of symbolic associations, and maintenance of multiple representations that support pre-algebraic thinking.

# Method

## Participants

Participants with CAL/WP-MD were selected from 1917 students 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). Before recruitment, we received Institutional Review Board approval. All teachers signed consent forms, and all students had signed consent forms from their parent/guardian. Some teachers participated in more than one study cohort, hence the discrepancy between number of classrooms and teachers. Stratifying by school, teachers (and their classrooms) were randomly assigned to three conditions: ~37.5% to CAL intervention (701 students); ~37.5% to WP intervention (705 students); and ~25% to control (CON; 511 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. The present study is a quasi-experiment due to a lack of complete randomness for all teachers (i.e., 26 teachers had participated in a previous cohort and retained those assignments; eight teachers received a condition to which they were not assigned; and two teachers fell into both categories). The parent report (Fuchs, Powell, et al., in press) indicated results were not substantively altered when only those who were newly and appropriately randomly assigned were evaluated.

The present analysis focused only on CAL/WP-MD students, whom we identified based on whole-class screening with the 1917 students on CAL and WP measures (see Measures). The benchmark for low performance in this report was scoring <7 on the CAL screening measure, <6 on the WP screening measure, and scoring above the 9th percentile on at least on subtest of the two-subtest Wechsler Abbreviated Scale of Intelligence (Psychological Corporation, 1999). These benchmarks for low performance were first calculated empirically using data from a pilot study where we identified pretest cut-off scores predictive of low performance on CAL and WP outcome measures. For the current report, we adjusted these benchmarks downward so that only students below the 40th percentile in both areas (i.e., CAL and WP) were included. With these cut-off scores, we identified 265 students from 110 second-grade classrooms taught by 83 teachers in 25 schools. Students were assigned to 1 of 3 tutoring conditions: 72 in Cohort 1 (27 in CAL; 24 in WP; 21 in CON); 56 in Cohort 2 (21 in CAL; 19 in WP; 16 in CON); 77 in Cohort 3 (29 in CAL; 33 in WP; 15 in CON); and 60 in Cohort 4 (19 in CAL; 15 in WP; 26 in CON). Screening measures correlated moderately at $r = .44$ in the entire screened sample and, according to their screening scores, the 265 students comprised the lowest 14% of the entire screened sample; the mean scores on the screening measures were obtained by only approximately

20% of the entire screened sample. Similarly, the mean standard score of CAL/WP-MD students was 85.88 on Wide Range Achievement Test-3-Arithmetic (Wilkinson, 1993). Mean standard score on KeyMath Problem Solving (Connolly, 1998) was 97.38, although this measure has a floor effect at the beginning of second grade, which makes it difficult for students to register low standard scores. Scores in the range noted above indicate high risk for poor outcomes, based on a previous study with a similar sample (Fuchs, Zumeta, et al., 2010). The mean standard score of CAL/WP-MD students was 85.88 on Wide Range Achievement Test Arithmetic (Wilkinson, 1993) and 97.38 on KeyMath Problem Solving (Connolly, 1998). (Note that KeyMath has a floor at beginning of second grade that makes it difficult for students to register low standard scores.) For the 265 CAL/WP-MD students, 96 were in CAL (36%), 91 in WP (34%), and 78 in CON (29%). Their teachers were similarly distributed (34.1%, 36.5%, and 29.1%, respectively), reflecting the design of the parent study. See Table 1 for demographic information by tutoring condition. Students' mean age was 7.57 years (*SD* = 0.40). For a breakdown of demographic information by cohort, contact the first author.

### Measures

**Screening—**We used three measures to determine high-risk status. We administered four subtests of the *Second-Grade Calculations Battery* (SGCB; Fuchs, Hamlett, & Powell, 2003): Sums to 12, Sums to 18, Minuends to 12, and Minuends to 18. All four subtests require single-digit addition or subtraction. Students have 1 min to complete up to 25 problems. Score is the number of correct answers. Alpha ranged from .85–.93. We used Sums to 12 as the CAL screening measure for determining high-risk students. We used the other three subtests to index calculation outcomes (see below). *Story Problems* (Jordan & Hanich, 2000) comprises 14 additive word problems. Problems represent combine, compare, and change word-problem types and require single-digit addition or subtraction for solution. The test is untimed. The tester reads each word problem aloud as the students follow along on paper, and then the tester provides time for students to write answers and do any word-problem work. A student's score is the number of correct answers. Alpha was .87. The *WASI* (Psychological Corporation, 1999) is an individually administered measure of general cognitive ability with two subtests. Vocabulary measures word knowledge and expressive language 4 picture items and 37 vocabulary words. Students define words until the end of the test or until reaching a ceiling of five consecutive errors. As reported by Psychological Corporation (1999), split-half reliability is .86. Matrix Reasoning measures nonverbal reasoning with pattern completion, classification, analogy, and serial reasoning tasks. Students choose the best of five choices to complete a visual patterns for 35 items or until reaching a ceiling of four consecutive errors or four errors over five consecutive items. Reliability is .94. The two subtest scores combine to yield an Estimated Full Scale IQ score with reliability of .92.

**Calculations—**We measured proximal single-digit calculations outcomes with the three non-screening subtests of the *SGCB* except (Fuchs et al., 2003). The outcome was the sample-based average of these three measures. To measure proximal double-digit calculations, we administered *Two-Digit Addition* and *Two-Digit Subtraction* (Fuchs et al., 2003). For each two-digit test, students have 3 min to complete up to 20 two-digit plus/

minus two-digit problems. Alphas for Two-Digit Addition and Two-Digit Subtraction were .
96 and .87, respectively. The outcome used was the sample-based average of these two
measures. We considered single-digit and double-digit calculations separately, given the
overall low pretest skill of these at-risk students.

We measured distal effects with *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. On the WRAT, students work on 40
calculations of increasing difficulty including addition, subtraction, multiplication, and
division of whole numbers and fractions. Students who do not meet a basal on the written
arithmetic also complete 15 oral arithmetic problems. Score is the number of correct
answers. Alpha was .93. On KM Addition, students answer 6 problems with pictures and 12
problems in a written format. Problems range from one-digit addition to addition with
decimals, fractions, and negative numbers. KM Subtraction functions similarly with 6
problems presented with graphics and 12 written problems of one-, two-, and three-digit
subtraction with whole numbers, fractions, decimals, and negative numbers. Both KM
subtests are scored by correct answers. Alphas for the two KM subtests were .84 and .81,
respectively. Within this sample, analysis indicated a single-factor solution across the three
tests (WRAT Arithmetic, KM Addition, and KM Subtraction) was a good fit to the posttest
data, and therefore used this factor score as the distal calculations outcome.

**Word problems—**We used two outcomes to measure word-problem performance: an
experimental measure of proximal effects and a factor score of commercial measures for
measuring distal effects. The proximal measure and distal composite correlated .52. *Second-
Grade Story Problems* (Fuchs, Seethaler, et al., 2008), the proximal measure, includes 18
word problems (never used for instruction) representing combine, compare, and change
problems types, with missing information in all three positions of the problem schema, with
and without irrelevant information, with and without charts or graphs, and with and without
multiple steps. Solutions require 1-digit addition or subtraction. Five items involve combine
problems, six involve compare problems, six involve change problems, and one item is a
two-step problem involving both combine and compare. In groups, the tester reads each
problem aloud as students follow along on paper. Students have 1 min to write a constructed
response before the tester reads the next problem. Students earn credit for correct word-
problem calculation (i.e., math answer) and label to reflect student's processing of the
problem statement and understanding of the problem's focus. Alpha was .88.

We measured distal effects with *KM Problem Solving* (Connolly, 1998) and the *Iowa Test of
Basic Skills - Data Interpretation and Problem Solving* (ITBS; Hoover, Hieronymous,
Dunbar, & Frisbie, 1993). KM Problem Solving includes 18 word problems of increasing
difficulty, which involve all four operations on word problems with transparent solution
strategies; non-routine problems without clear solution strategies; and items requiring
students to demonstrate comprehension of a word problem without solving it.
Administration is individual, and items are presented orally with accompanying pictures.
Students answer with an oral response. Testing is discontinued after three consecutive
errors. Score is the number of correct multiple-choice answers. Alpha was .74. ITBS was
administered in groups. Students work on eight word problems presented orally and 22

written problems. Problems include one- and two-digit calculations with and without irrelevant information, with and without charts or graphs, and with and without multiple steps. The tester reads all problems aloud, and students respond using a multiple-choice format. Alpha was .81. These two measures (i.e., KM Problem Solving and ITBS) correlated $r = .43$ and were combined into a sample-based composite score.

**Pre-algebraic knowledge**—Pre-algebraic knowledge outcomes included proximal and distal measures administered only at posttest. 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, Powell, et al., 2009), students solve equations ($a + b = c$ or $d - e = f$) that vary the position of "$x$" across all three positions. The score is the number of correctly solved equations. Alpha on this sample was .91. With Number Sentences (Fuchs, Powell, et al., 2009), the tester reads eight word problems aloud; students have 30 sec to write the mathematical equation representing the problem structure (students do not find solutions). The score is the number of correct equations. Alpha was .84. These measures (i.e., Find X and Number Sentences) correlated $r = .47$ and were combined into a composite score.

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 student's responsiveness to instruction on finding the missing variable in addition expressions (Skill A; e.g., $x + 5 = 11$ or $6 + x = 10$), simple multiplication expressions (Skill B; e.g., $3x = 9$), and equations with two missing variables (Skill C; e.g., $x + 2 = y - 1$; $y = 9$). 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 indicates students never mastered any skill; 21 indicates students mastered each skill on the pretest and got each generalization problem correct). Alpha was .84. The outcome was a sample-based *z*-score.

In Cohorts 2 through 4, a composite was generated across DA and the *Test of Pre-Algebraic Reasoning* (TPAR; Fuchs, Seethaler, & Powell, 2009), which comprises two types of problems. The first problem types (20 items) are mathematical equivalence statements with letters (i.e., *x* and *y*) representing missing quantities: 18 in nonstandard format (e.g., $y + 4 = 9$); two in standard format (i.e., $1 + 5 = x$). The next problem types (6 items) are function tables, each of which shows a two-column table. The first column shows a variable (*x* or *y*); the second shows a function involving that variable; each row shows a value for the variable and the resulting value for the function. In one row, the value of the function is empty and shaded; the task is to complete that row. 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 with mathematical equations and as much time as needed (until all but two students finish) to complete the second problem type with function tables. 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 was .88. In Cohorts 2 through 4, the score was a sample-based *z*-score composite of DA and TPAR.

### Study Conditions

Students participated in two-tiered CAL intervention, two-tiered WP intervention, or business-as-usual control. Tier 1 comprised 34 whole-class lessons with two lessons per week over a 17-week period. Each lesson lasted approximately 40 to 45 min. Research assistants (RAs) delivered all lessons in the students' classrooms. Tier 2 comprised 39 small-group tutoring lessons. Tutoring started at the beginning of weeks 4 or 5 of the Tier 1 intervention and lasted for 13 weeks with 3 lessons a week. RAs provided tutoring in groups of 2 to 3 students, and each lesson lasted 25 to 30 min. Tier 2 tutoring took place outside of the student's classroom (i.e., library, conference room, hallway), and Tier 2 tutoring did not occur at the same time in the school day as Tier 1 intervention.

### Instructional Framework to Support Connections with Pre-Algebra

We relied on Pillay et al.'s (1998) model to design the CAL-RTI instructional whole-class and tutoring programs and to design the WP-RTI instructional whole-class and tutoring programs in ways that potentially support the transition from arithmetic to algebra. In this section, we provide an overview of the guiding frameworks. For information on the specifics of the programs, consult Fuchs et al. (in press). Contact lynn.a.davies@vanderbilt.edu for program manuals under the title *Math Wise* for CAL whole-class instruction and tutoring and *Pirate Math* for WP whole-class instruction and tutoring (*Math Wise* because the theme is owls, with allusions to becoming "wiser" in mathematics by using appropriate calculation strategies; *Pirate Math* because the theme is pirates, with allusions to finding X, the unknown, in word problems, just as X marks the treasure on pirate maps.)

**CAL framework—**CAL whole-class intervention and small-group tutoring incorporated two areas of emphasis for developing conceptual and procedural competence with one- and two-digit addition and subtraction calculations (e.g., Fuchs et al., 2013; Fuson & Kwon, 1992; Geary et al., 2008; Groen & Resnick, 1977; LeFevre & Morris, 1999; Siegler & Shrager, 1984). We focused on interconnected knowledge about number (e.g., cardinality, inverse relation between addition and subtraction, commutative property). Students used connecting cubes to explore how a target number (e.g., 7) can be partitioned in different ways. For example, 7 can be made of 3 and 4 ($3 + 4 = 7$), and 2 can be taken away from 7 for a difference of 5 ($7 - 2 = 5$). Students focused on part-part-whole knowledge related to number families (e.g., $2 + 5 = 7$, $5 + 2 = 7$, $7 - 5 = 2$, $7 - 2 = 5$) and grouped families using manipulatives or graphics on paper to demonstrate how and why problems constitute a number family. Students also used number families to explore the inverse relation between addition and subtraction. We also included a focus on place value concepts of tens and ones. Students practiced counting by tens, learned about the meaning of zero in the tens and ones places, and used and regrouped Base-10 manipulatives to represent one- and two-digit numbers.

The other area for emphasis in CAL-RTI was practice. Students learned counting strategies for solving one-digit problems and two- plus/minus one-digit problems without regrouping. Practice required students to generate many correct responses to such problems to help them form long-term representations to support retrieval. Practice helped build fluency, which, in turn, helped improve double-digit addition and subtraction. Research assistants also explicitly taught efficient procedures for identifying when regrouping was required in addition and subtraction problems and how regrouping differed in addition and subtraction.

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 (problems where both operands are 1-digit or one operand is 1 digit and the other is 2 digits but regrouping is not required); (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 (although cumulative review was also integrated throughout the first five units).

**WP framework—**WP intervention incorporated Kintsch and colleagues' framework in which WP solving is thought to be an interaction between problem solving and language comprehension (Cummins, Kintsch, Reusser, & Weimer, 1988; Kintsch & Greeno, 1985; Nathan, Kintsch, & Young, 1992). Their model of WP solving suggests general features of understanding text apply across all WPs but that WPs differ by problem type. Students can learn to apply a problem-type schema to assist with solving specific WPs, and at second grade, the three additive schemas include combine, compare, and change (Riley & Greeno, 1988).

At the beginning of instruction, teachers taught a general WP attack strategy (i.e., RUN) to use on all WPs: Read the problem, Underline the labels, and Name the problem type (i.e., schema). After learning the attack strategy, we focused on teaching students how to recognize the underlying structure of a WP and use one of three schemas to solve a WP. Initially, teachers introduced each schema by presenting a WP without missing information to enable students to focus on the WP structure. Teachers asked students to focus on the action or lack-of-action in the word problem as well as the content of the word-problem question. We used graphics to demonstrate the relations underlying each schema, and we provided students with the opportunity to place known information into the graphics. After learning about the underlying structure of a WP, teachers taught students "meta-equations" that represented each schema and could be used to plug in known information from the WP. The meta-equation for combine problems was $P1 + P2 = T$ (i.e., *part 1* plus *part 2* equals the *total*). The meta-equation for compare problems was $B - s = D$ (i.e., the *bigger* amount minus the *smaller* amount equals the *difference*), and the change meta-equation was $St +/- C = E$ (i.e., the *start* amount has an increased or decreased *change* for a new *end* amount). After determining the word-problem schema, students wrote the meta-equation and filled in known word-problem information under specific parts of the meta-equation. Teachers also taught students to represent missing WP information with an "X."

We also provided explicit instruction on the language comprehension necessary for combine, compare, and change problem types. As in Jitendra (2007), we focused on helping students develop understanding of relational terms for compare WPs (e.g., *more/less; older/ younger; hotter/colder*). Also to extend Jitendra (2007), we also focused on relational terms in static (i.e., combine and compare) versus change WPs (e.g., *Claire had 7 fewer candies than Winston* versus *Claire had 7 candies and gave 2 to Winston*) and vocabulary related to quantities (e.g., *amount* refers to a quantity) and superordinate levels (e.g., *4 cows and 9 horses* equals *13 animals*) for combine WPs. Note, however, that we did not teach students "key words," which consensus suggests is not a helpful WP solving strategy (Clement & Bernhard, 2005). Instead, we focused on the deeper understanding of specific words and phrases for WP interpretation.

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 (although cumulative review was also integrated throughout the first four units). 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—**Both CAL and WP intervention linked to pre-algebraic knowledge, as per Pillay et al. (1998), in two ways. First, in both CAL and WP intervention, we focused on equal-sign understanding. Specifically, teachers taught students to understand the equal sign as a relational symbol, indicating the two expressions on either side of the equal sign balanced (Jacobs, Franke, Carpenter, Levi, & Battery, 2007). While using terminology (e.g., "same as") to describe the equal sign improved student understanding of the equal sign (Baroody & Ginsburg, 1983; Blanton & Kaput, 2005), explicit instruction on the location or meaning of the equal sign also improves student understanding (McNeil & Alibali, 2005; Rittle-Johnson & Alibali, 1999). As part of WP intervention, Powell and Fuchs (2010) determined that third-grade students with MD who had equal-sign instruction embedded within WP tutoring performed better than students without embedded equal-sign instruction on solving equations and some types of WPs.

Second, we used "meta-equations" to help organize word-problem work by schema. For combine WPs, where two Parts combine for a Total, the meta-equation was P1 + P2 = T. For compare WPs, where we compare the Bigger number to the smaller number for a Difference, we used B − s = D. For change WPs, we used St +/− C = E to represent a Starting amount that increases or decreases (i.e., Changes) for a new End amount. Students learned to identify the WP type and write the corresponding meta-equation. Then, using the meta-equation as a guide, students filled in the appropriate numbers from the WP and placed an X in the equation to represent the missing information (i.e., WP answer). Teachers taught specific strategies for solving for X, which was shown to increase the pre-algebraic reasoning of second-grade students (Fuchs, Zumeta, et al., 2010). As the WP intervention provided this additional pre-algebraic component (i.e., using meta-equations), we expected WP students to demonstrate stronger pre-algebraic reasoning than CAL students.

**Teachers' instruction—**The primary mathematics curriculum of the school district was *Houghton Mifflin Math* (Greenes et al., 2005). All teachers used this curriculum to guide mathematics instruction in second grade. The content of the curriculum matched the CAL intervention in terms of teaching one- and two-digit addition and subtraction. The content also matched the WP intervention by focusing on combine, compare, and change word-problem types. In this way, CON students received CAL and WP instruction similar to that of the two interventions. See Fuchs et al. (in press) for a discussion of the similarities and differences among CAL intervention, WP intervention, and the district math curriculum.

**Structure of CAL and WP whole-class intervention—**Both CAL and WP whole-class instruction occurred for 17 weeks with two 40–45 minute lessons each week. In both *Math Wise* and *Pirate Math*, RAs provided explicit instruction. Each lesson began with an explanation of worked examples. Then, the RAs used guided practice to demonstrate strategies on partially worked and unworked problems. Next, students worked in pairs as the RA circulated and provided feedback as necessary. Finally, students participated in an independent practice activity related to the lesson's skill. Cumulative review was systematically incorporated throughout both CAL and WP interventions.

In CAL whole-class instruction, this general framework is captured in four activities per lesson. The first, the *Daily Lesson*, focuses on six to eight problems on the day's topic. This involves the teacher presenting worked examples and moving to partially worked and then unworked problems, with constant student participation. The second activity is *Partner Work*, in which paired students (a higher- and lower-performer) work in a structured manner on 12 problems on that day's topic, as coach and player. As validated with Peer-Assisted Learning Strategies (Fuchs et al., 1997), the higher-performing student begins as coach, who asks the player step-by-step questions to model the teacher's solution strategy and provides corrective feedback. This occurs on three problems. Then the player "thinks-aloud" the solution strategy on the next three problems, while the coach monitors and provides corrective feedback. Students then switch roles, so the lower-performer asks his/her partner to model the solution strategy on problems 7 through 9, while providing corrective feedback, and finally the higher performer talks aloud the final three problems. The teacher circulates while providing assistance. The third activity is *Time Owl* (beginning Lesson 15), which presents a page of different types of calculation problems. The teacher provides a directive (e.g., "solve addition problems that require regrouping"); then students have 1 min to find and complete that type of problem. At the end of 1 min, students switch Time Owls with partners and check answers. The last activity, *Individual Practice*, provides students with a practice sheet of 25 problems on the day's topic and cumulative review. At the end of 5 min, students switch papers to check answers. During the final three activities, students earn points for working with partners cooperatively and producing accurate work. The pair with the most points collects Math Wise folders.

WP whole-class instruction also includes four activities per lesson. The first is the *Teacher-Led Problem*, in which the teacher reviews and leads discussion about a problem from the previous day's lesson. Each student has a copy of that problem, which the teacher and students work through together. The second activity is the *Daily Lesson*, focused on that day's topic. This involves the teacher presenting worked examples and moving to partially-

worked and then to unworked problems, with constant student participation. The third activity is *Partner Work*, in which students work in pairs to solve two word problems on the day's topic. After ~8 min, the teacher shares answers with the class, against which the pair compares its work. The final activity is *Individual Practice*, in which students complete five find x problems and one word problem. As each student finishes, the teacher checks his/her work. Students earn points for each part of a correctly applied solution strategy and record scores on a "treasure map."

**Structure of CAL and WP tutoring—**An RA, different from the whole-class RA, conducted all tutoring lessons. Tutoring did not occur during the same time slot as whole-class instruction to ensure that students participated in both tiers of intervention. Tutoring content was similar to whole-class content, except that tutoring targeted the most difficult concepts from whole-class instruction, used manipulatives more frequently, employed additional scaffolding from the RA, taught additional strategies, and incorporated self-regulation with tangible reinforcements. Throughout each tutoring lesson, RAs asked questions to individual students or the small group to check for engagement and understanding. RAs set a timer at random intervals (as outlined in each lesson's guide) and awarded a check if all students in the group were following directions and working hard when the timer beeped. At the end of each lesson, students earned individual checks for correctly answering "bonus problems." Students did not know which problems worked during tutoring would end up as "bonus problems" until the end of each lesson. This encouraged students to do their best mathematics work throughout the lesson. Students colored an owl (for *Math Wise*) or a pirate-themed picture (for *Pirate Math)* for each checkmark earned during the lesson. After a student colored 16 pictures, which typically took 3 to 4 tutoring lessons to complete, the student selected a small prize out of a treasure box.

Six activities occurred during each CAL tutoring lesson. Fist, students participated in *Number Combination Flash Cards*. For Lessons 1–13, students worked on addition flash cards. Starting with Lesson 14, subtraction flash cards were mixed in with the addition cards. For 1 minute, the RA showed cards to the students in a round robin format. If a student answered incorrectly, the RA encouraged the student to use a counting strategy to answer the number combination correctly. At the end of 1 minute, the RA and students counted the correctly-answered cards. Then, the RA shuffled the cards, and the students had another 1 min to try to beat their score. The group then graphed their highest flash card score on a graph. The second activity each lesson was the *Number Combinations Lesson.* The RA reviewed number knowledge content (e.g., cardinality, mental representation of the number line, relationship between addition and subtraction, commutative property of addition) and efficient counting strategies (i.e., min strategy for addition; missing addend strategy for subtraction). Lessons began with a focus on number combinations of +1/−1, moving to +0/−0, and then doubles. After doubles, lessons focused on number combinations in sets (e.g., all combinations with a sum of 5 or 5 as the minuend). The RA encouraged students to use manipulatives, such as snap cubes, to understand sets of combinations. The third activity each lesson was *Owl Belly*, an activity named by the students. Each student had a worksheet with eight owls, and each owl had a number on its belly (e.g., 11). The tutor read a number

combination aloud (e.g., "four plus seven"), and the students have the color the owl with the correct answer. RAs presented only six problems, and students had to write one addition and one subtraction number combination for the two uncolored owl bellies. The fourth activity each lesson was *Double-Digit Flash Cards.* Similar to the first flash card activity, students responded in a round robin format for 1 minute and answered whether the problem on the flash card required addition or subtraction (Lessons 1–21) or whether the problem required regrouping and why (Lessons 22–39). The fifth activity was *Double-Digit Lesson*. Here, the RA reviewed stories and rhymes taught in CAL whole-class intervention, which taught place value and regrouping concepts. Students completed six double-digit problems with Base-10 blocks. The sixth activity each lesson was *Paper-and-Pencil Review*. Students had 1 minute to independently solve 10 one-digit problems and 2 minutes to independently solve 5 two-digit problems. RAs provided corrective feedback as necessary.

Four activities occurred each WP lesson with the first activity, *Game,* differing by lesson topic. Students played the Find X Game (solving for X in equations), the Total Game (filling in the Total meta-equation for orally-presented problems), or the Difference Game (underlining and drawing inequality symbols between two compared amounts in a word problem). During the *Daily Lesson*, RAs provided scaffolded schema instruction as students completed three to four word problems. Most word problems aligned with the unit's focus, but previously-learned word problem types were incorporated to help students practice discrimination among problem types and to provide cumulative review. With the third activity, *Sorting Game*, the RA read word problems and students respond, round-robin style, for 2 min as to the problem type of the word problem. At the end of 2 min, the RA provided corrective feedback for incorrectly-identified word problems. The fourth activity each lesson was *Paper-and-Pencil Review*. Students worked individually for 2 min on solving nine Find X problems and for 2 min on solving one word problem. RAs provided feedback as necessary.

**RAs**—Across the four cohorts, 39 RAs provided tutoring to 143 groups of students in groups of two or three students with several "groups" of one created after students left the tutoring project. A typical RA was one or two years beyond undergraduate education and studying for a graduate degree in an education-related field. To control for quality of RAs, every RA taught or tutored in both CAL and WP intervention. We guarded against contagion of conditions by coding lesson guides and materials by condition; reviewing the differences between the two conditions in weekly meetings; and providing corrective feedback during in-school observations and digital recordings of small-group tutoring lessons.

To prepare RAs for tutoring duties, we conducted an introductory meeting during which we discussed the ethics of conducting school research. Then, during two six-hour sessions, RAs learned how to implement CAL and WP whole-class teaching or small-group tutoring. We introduced RAs to the goals of the project, and we provided instruction and demonstration on implementing the lessons. RAs were paired together to practice delivery of lessons, and before providing tutoring in schools, RAs had to demonstrate proficiency with a project coordinator by delivering a lesson. RAs who achieved 95% fidelity against the lesson's fidelity checklist were judged ready for school tutoring. RAs who scored below 95%

received coaching from the project coordinator, practiced lesson delivery, and then were re-assessed at a later time on another lesson. During weekly meetings held throughout the year, the first author, project coordinators and RA-tutors discussed upcoming tutoring lessons and any tutoring difficulties.

**Fidelity—**To assess fidelity, all sessions were digitally audio-recorded, and we conducted in-school observations of each RA. Prior to the first tutoring session, we created checklists of essential information for each lesson within CAL intervention and WP intervention. At the end of each school year, RAs independently listened to a random sample of ~20% of tutoring tapes, representing conditions, RAs, tutoring groups, and lesson types equitably, while completing the checklists to identify the percentage of points addressed. 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, RA was the unit of analysis; $t$ tests are for dependent samples because each RA taught in both CAL and WP conditions.)

### Procedure

Over the four cohorts, we administered the screening measures in September and the pretest measures in October. RAs delivered whole-class instruction from November to March and tutoring from December to March. We posttested all students in late March. RAs independently entered responses on 100% of the test protocols for every screening and outcome measure on an item-by-item into two separate databases. We compared and rectified the discrepancies between the two databases to reflect the student's original response.

### Analysis Plan

Multilevel modeling was utilized for primary analyses within SAS PROC MIXED (SAS Institute, 2009), with restricted maximum likelihood estimation, and reporting standard model fit indices (deviance or −2 log likelihood; Bayesian Information Criteria or BIC). We followed a model-building approach, where we first established the clustering effects, or intraclass correlation (ICC), of both teacher and school in unconditional models (Luke, 2002; Sullivan, Dukes, & Losina, 1999). Where the random effect of school was not contributory, we eliminated this term from the random portion of the model. Next, student-level fixed effects were added to the model, primarily the pretest variable associated with the dependent variable being analyzed. For CAL outcomes, the pretest covariates corresponded to the outcomes: single-digit SGCB, double-digit SGCB, and WRAT (for the standardized outcome). For WP, the pretest factor score was computed from SGSP and KM Problem Solving. Covariates for algebra outcomes were based on the strength of their relations with pretest variables: Proximal algebra was used for the word-problems factor score; distal algebra for the word-problem factor score and WRAT performance. Additional student-level covariates included sex, race/ethnicity, reduced/free lunch (RFL) status, special education status, and English as Second Language (ESL) status, where these were related to outcomes. Treatment condition was then added as a teacher-level effect; cohort was also considered a teacher-level effect. Interactions among these terms, and interaction terms across tiers were also considered and discussed below where relevant.

To represent the size of the effect of treatment in addition to statistical significance, we used two measures. The primary standardized effect size (ES) was computed using parameter estimate differences between treatment groups, divided by the square root of the student-level residual variance, when treatment was the sole factor in the model. We also computed a proportional reduction in variance (PRV; Peugh, 2010; Singer & Willett, 2003) that is based on how well the final model does in reducing between-teacher variance, relative to a model with all other predictors except treatment; such values do not represent the absolute amount of variance explained in the outcome (Ma & Ma, 2008).

## Results

Table 1 shows demographic information by tutoring condition. There were no significant differences on race/ethnicity, RFL status, special education status, ESL status, or retained status. There was a significant difference for gender, $\chi^2(2, N = 265) = 6.21$, $p = .045$. Table 2 shows means and *SD*s for each screening and outcome measure. For a breakdown of screening and outcomes measures by cohort, contact the author. There was no significant difference as a function of treatment condition on any pretest measure. Table 3 shows results for the two proximal calculation outcomes; Table 4 for the distal calculation outcome; Table 5 for the proximal and distal word-problem outcomes; and Table 6 for the proximal and distal pre-algebraic knowledge outcomes. The top portion of each table shows results of the unconditional models, along with model fits. The bottom shows results for the final model including model fit, effects for treatment, cohort, pretest, and demographically relevant covariates.

### Proximal Calculation Effects

For single-digit outcomes, the unconditional model suggested a random effect only for teacher, with a strong ICC of .44. The final model included teacher as a random effect; for fixed effects, at Level 1 (L1), once pretest was in the model, no other predictors at this level were significant. Level 2 (L2) teacher-level predictors were treatment and cohort; there was no interaction between pretest and treatment. In the final model, there were significant main effects for pretest ($p < .001$, with higher outcomes associated with higher pretest performance) and treatment condition, $F(2, 89.1) = 30.56$, $p < .001$. Cohort did not interact with treatment and was not a significant effect alone, $F(3, 148) = 2.36$, $p = .074$. Follow up to the treatment effect indicated CAL outperformed WP ($p < .001$; ES = 0.95) and CON ($p < .001$; ES = 1.67), and WP outperformed CON ($p < .001$; ES = 0.71). Adding treatment condition resulted in a PRV of 61% across teachers, relative to the prior model with all predictors except treatment.

For double-digit calculations, the unconditional model also included only teacher as a random effect, with an ICC similar to single-digit calculations (.45). The final model included both teacher and school as random effects; fixed effects included pretest (at L1) and cohort and treatment (at L2). In the final model, there were no significant interactions. There were positive effects for pretest ($p = .003$), for cohort, $F(3, 50.6) = 3.37$, $p = .026$, and for treatment condition, $F(2, 64.9) = 35.65$, $p < .001$. CAL outperformed WP ($p < .001$; ES = 1.21) and CON ($p < .001$; ES = 1.67), but WP was comparable to CON ($p = .089$; ES =

0.45). For double-digit calculations, PRV for treatment condition was 73% across teachers and schools. Thus, results were directionally similar to those found for single-digit calculations, except WP was comparable to CON on double-digit problems.

### Distal Calculation Effects

On the distal calculation outcome, the unconditional model revealed an ICC of .03 for school and .20 for teacher. The final model included both teacher and school as random effects; fixed effects were pretest, age, and special education status (at L1) and cohort and treatment (at L2). In the final model, there were no significant interactions. There were positive effects for pretest ($p < .001$), age ($p < .001$), special education ($p = .008$), and treatment condition, $F(2, 66.6) = 34.68$, $p < .001$, but not cohort ($p = .713$). CAL outperformed WP ($p < .001$; ES = 0.72) and CON ($p < .001$; ES = 1.19); WP also outperformed CON ($p = .005$; ES = 0.48). Adding treatment resulted in a PRV of 73% across teachers and schools for distal calculations.

### Proximal Word-Problem Effects

For the unconditional model, the ICC for teacher was .43; school was non-contributory. The final model included both teacher and school as random effects; fixed effects were pretest, age, special education status (at L1), and cohort, treatment, as well as the interaction between cohort and treatment (at L2). In the final model, there was a significant interaction of cohort and treatment, $F(6, 73.2) = 5.77$, $p < .001$, that suggested smaller differences between WP and the other conditions in the first two cohorts relative to the third and fourth cohorts. Even in the context of this interaction, there were positive effects for pretest ($p < .001$), special education ($p = .038$), cohort ($p = .002$), and treatment condition, $F(2, 40.7) = 44.77$, $p < .001$. WP outperformed CAL ($p < .001$; ES = 1.47) and CON ($p < .001$; ES = 1.31), but CAL and CON were comparable ($p = .152$; ES = −0.16). In Cohorts 3 and 4, the pattern of treatment conditions fit the overall pattern. However, in Cohort 1, students WP outperformed CAL but not CON; in Cohort 2, there were no significant differences. By contrast, in Cohorts 1 and 2, WP outperformed CAL and CON, but CAL and CON were comparable. The PRV for treatment was 88% across teachers and schools for proximal word problems.

### Distal Word-Problem Effects

On distal WP outcomes, the unconditional model revealed an ICC of .06 for school; the ICC for teacher was non-contributory. For completeness, the remaining steps were followed and presented here, although given little variability in outcomes across teachers, treatment effects (assigned at the teacher level) are unlikely. The final model included school as a random effect; fixed effects were pretest and age (at L1) and cohort and treatment (at L2). In the final model, there were no significant interactions. There were positive effects for pretest ($p < .001$) and age ($p < .001$), but not for cohort ($p = .304$) or treatment ($p = .871$). The ES for WP versus CAL (0.05) and to CON (ES = 0.15) were negligible, as was the ES for CAL versus CON (0.09). The PRV for the addition of treatment was < 1%.

### Proximal Pre-Algebraic Knowledge Effects

For the unconditional model, the ICC for teacher was .37; school was non-contributory. The final model included teacher as a random effect, and fixed effects were pretest and special education status (at L1) and cohort and treatment (at L2). There were positive effects for pretest ($p < .001$), special education ($p = .006$), and treatment condition, $F(2, 81.1) = 58.38$, $p < .001$, but not for cohort ($p = .538$). WP outperformed CAL ($p < .001$; ES = 1.52) and CON ($p < .001$; ES = 1.67), but CAL and CON were comparable ($p = .802$; ES = 0.16). The PRV for treatment condition was 88% across teachers.

### Distal Pre-Algebraic Knowledge Effects

For the unconditional model, the ICC for teacher was .11; for school, the ICC was .01. The final model included teacher as a random effect, and fixed effects were pretests (both word-problem composite and WRAT-3 computations), special education status (at L1), cohort, and treatment (at L2). There were positive effects for word-problem pretest ($p < .001$), WRAT-3 ($p < .013$), special education ($p = .005$), and treatment condition, $F(2, 68) = 3.38$, $p = .040$, but not for cohort ($p = .099$). WP outperformed CON ($p = .016$; ES = 0.52). However, WP was comparable to CAL ($p = .063$), with an ES of 0.22. CAL and CON were also comparable ($p = .470$; ES = 0.25). Adding treatment condition results in a PRV of 19% across teachers and schools relative to the prior model.

## Discussion

In the present study, we focused on the efficacy of multi-tier mathematics RTI systems in calculations and word problems for second-grade students at-risk for CAL-WP/MD. We compared students at-risk for CAL-WP/MD who received two tiers of support (i.e., whole-class instruction and small-group tutoring) in either calculations or word problems to each other and to a group of business-as-usual control students who were also at-risk for CAL-WP/MD. This was our major contrast, but we also included the business-as-usual condition to control for history and maturation effects and to determine whether students in the RTI conditions made more progress than would have occurred without RTI.

On both proximal and distal calculation outcomes, CAL students outperformed WP and control (CON) students with ESs ranging from 0.72 to 1.67. Our CAL-RTI interventions, focused on establishing conceptual understanding and building fluency with single- and double-digit addition and subtraction calculations, proved effective. Interestingly, on the distal calculation outcome, WP students outperformed CON students (ES = 0.48), even as CAL students demonstrated superior performance over WP students (ES = 0.72). This result is not surprising as WP students performed calculations on word problems during every WP lesson. Our hypothesis related to calculation outcomes, that CAL-RTI would promote superior calculation outcomes, was corroborated. Because students with MD often demonstrate a lack of fluency with number combinations (Baroody, Bajwa, & Eiland, 2009) and given that difficulty with number combinations often contributes to difficulty with solving more complex computation problems (Chong & Siegel, 2008) and higher-level mathematics problems in the areas of rational numbers, algebra, and problem solving (Wei,

Lenz, & Blackorby, 2012), this finding is also important: The CAL-RTI approach shows promise for establishing strong calculation skill for students with CAL/WP-MD.

On proximal word-problem measures, WP students demonstrated significant growth over CAL and CON students with ESs of 1.31 to 1.47. Interestingly, WP results from Cohorts 1 and 2 were not as strong as results from Cohorts 3 and 4. Between Cohorts 2 and 3, we designed a language component for both whole-class and small-group WP instruction, and this component was embedded within WP instruction for Cohorts 3 and 4. At the whole-class tier, RAs presented students with "compare sentences," and the RA and students worked together to identify a "compare word" to designate which item/person has the bigger or smaller amount. To reduce demands on working memory, students were taught to place an inequality symbol (i.e., < or >) above the compare word. The RA then worked with the class to write a new compare sentence using an opposite compare word, for which the relationship between the bigger and smaller amounts was preserved. Within small-group tutoring, the language component included working on two compare sentences, placing the inequality symbol over the compare word, and writing B (for Bigger amount) and s (for smaller amount) over the items/persons compared. Students then filled in the Difference meta-equation (i.e., B − s = D) and placed X in place of the missing information. While difficult to pinpoint the exact cause for better WP performance from Cohorts 3 and 4, the added focus on the language used in word problems is a likely contributor to better word-problem understanding, by problem type, and/or increased understanding of the relationship between numbers in word problems. The language component may be especially important for students with CAL/WP-MD because these students often experience comorbid language and reading difficulties (Fuchs & Fuchs, 2002). Even so, on distal word-problem measures, there were no significant differences among CAL/WP-MD students in the three conditions, and this result mirrors previous word-problem studies using commercial word-problem measures as outcomes (Fuchs, Zumeta, et al., 2010). And it is important to note that the proximal measure shared important design elements with commercial word-problem tests: It mixed different problem types, while requiring transfer to untaught program components. Therefore, with strong word-problem effects on this proximal measure, we corroborate our hypothesis that WP-RTI positively increases word-problem outcomes. Even though successful word-problem performance may not necessarily translate to success with real-world problem solving, word problems are important because they represent a major emphasis in every strand of the mathematics curriculum (i.e., solving word problems using fractions, geometry, measurement, algebra, etc.) and because word-problem difficulty has implications for both successful school and occupational success (Murnane, Willett, Braatz, & Duhaldeborde, 2001; Parsons & Bynner, 1997). With a WP-RTI approach, the WP skill of students with CAL/WP-MD can be improved.

We were also interested in how CAL and WP intervention transferred to performance on pre-algebraic measures. Confirming our hypothesis that WP-RTI would provide a stronger route than CAL-RTI to pre-algebraic reasoning, WP students outperformed CAL and CON students on proximal measures of pre-algebraic reasoning with ESs of 1.52 to 1.67. Moreover, CAL and CON students performed similarly. On distal measures of pre-algebraic reasoning, WP demonstrated a significant gain over CON students (ES = 0.52) and a marginally significant gain over CAL students (ES = 0.22). Similar to proximal measures of

pre-algebraic reasoning, CAL and CON students performed comparably. Our results therefore indicate WP-RTI is more influential than CAL-RTI in improving the pre-algebraic reasoning of students with CAL/WP-MD. This finding adds to Fuchs et al.'s (2012) correlational study, in which the connection to algebra for WPs was stronger than CAL.

Kieran's (1990) developmental progression on algebraic competence suggests arithmetic competence is a foundational skill necessary for algebra performance. Kieran states that part of arithmetic competence involves understanding symbols and solving equations. Even though both CAL and WP interventions focused on using operator symbols (e.g., +, −, =) and solving equations, we have two potential explanations for the superior pre-algebraic performance of students in the WP-RTI intervention. First, the WP-RTI interventions required students to develop conceptual understanding attached to symbols presented through real-life scenarios (i.e., word problems). For example, students interpret "Annabelle lost 8 flowers" as a subtraction expression, and "Lincoln bought some more candy" as an addition expression. By learning problems types by schema, students worked on understanding the underlying structure of an addition or subtraction number sentence. Second, the WP-RTI interventions taught students to set-up and solve equations where a sum or difference was not always the required computation. For example, CAL students always solved equations such as $14 - 9 = \_\_$ or $3 + 8 = \_\_$. WP students solved equations such as $14 - 9 = X$, but also solved equations such as $3 + X = 11$ or $X - 4 = 7$. To solve these latter equations, students had to evaluate the missing part (i.e., X) and determine how to solve for X. Solving for X sometimes required not using the operator symbol in the equation (e.g., $X - 4 = 7$ is solved by adding 4 and 7 together). By solving equations with a missing variable, WP students worked on Kieran's second stage of algebraic competence where the focus is understanding the equal sign as relational and understanding the concept of a variable. These results have important implications for designing early intervention for students at-risk for long-term MD, including difficulty with algebra. Findings indicate that an instructional focus on word problems, which is deliberately designed to forge connections with pre-algebraic thinking (as in the present study), assists students in their transition to formal algebraic thinking. Finding this link with a sample of students with concurrent difficulty with calculations and word problems is notable.

### Conclusion

In sum, we learned that a multi-tier support system focused on calculations improves the calculation outcomes of students at-risk for CAL/WP-MD and that, to realize benefits on double-digit addition and subtraction problems, explicit attention to those concepts and algorithms is required. At the same time, we found no evidence that intervention on calculations transfers to word-problem outcomes, which corroborates prior research (e.g., Fuchs et al., 2011; Fuchs, Powell, et al., 2009). At the same time, we learned that a multi-tier support system focused on word problems improves word-problem performance of students at-risk for CAL/WP-MD. However, in contrast to CAL-RTI, for which we saw no indication of transfer to word-problem skill, we did see evidence of transfer from WP-RTI to single-digit (but not double-digit) calculation problems. Given that competence with calculations and word problems is necessary for success in adulthood, findings suggest efficiency for WP-RTI over CAL-RTI, even as results indicate that instruction on word problems, which

does not explicitly address calculations, does not transfer to improved performance on double-digit calculation problems that require regrouping.

Because the research base on pre-algebraic reasoning is emerging, we were also interested in learning whether there were differences between CAL and WP for improving pre-algebraic reasoning. WP-RTI improved the pre-algebraic skill of students at-risk for CAL/WP-MD significantly more than occurred in the CAL-RTI condition or in the control group. This indicates that components of WP instruction, as designed in the present study to support the connection with pre-algebraic thinking, contribute to successful pre-algebraic reasoning. As algebraic reasoning is an integral part of mathematics standards at the elementary, middle, and high school grades (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) and as many students are unprepared for algebra courses (NMAP, 2008), elementary school curricula need to emphasize algebraic reasoning. By using WP interventions focused on teaching the underlying structure of the problem, using equations to represent word-problem schema, embedding equal-sign instruction within WP intervention, and solving equations with a variable (e.g., X), the pre-algebraic competence of students with CAL/WP-MD improves, and this may lead to better success with mathematics, especially algebra, in later grades.

In these ways, this study's WP and CAL multi-level support systems enhanced the mathematics performance of high-risk students – those with concurrent difficulty with calculations and word problems. These students, who are already performing substantially below grade-level expectations, must demonstrate improvement in word problems and calculations in order to meet ambitious mathematics standards in late elementary, middle, and high school. By design, to assess transfer across mathematics domains, our CAL-RTI focused exclusively on teaching calculation skill, with strong effects on calculations. Our WP-RTI, which focused exclusively on word problems, provided greater cognitive demands by challenging students to recognize word problems by type, write number sentences to represent word-problem schema, and solve equations with an unknown value. If teachers must choose between providing WP or CAL intervention, given constraints on mathematics intervention time in most schools, study results favor WP-RTI: Whereas CAL-RTI only enhanced calculation skill, WP-RTI simultaneously improved students' competence with word problems, single-digit calculations, and higher-level problems related to pre-algebraic reasoning.

### Implications for Practice

This study therefore suggests that teachers, when addressing the needs of students with co-occurring calculation and word-problem difficulty, must allocate specific instructional focus to word problems, rather than assuming that improvement in calculations will translate into improvement with word problems. Moreover, WP-RTI significantly improved pre-algebraic reasoning moreso than did CAL-RTI on pre-algebraic reasoning. So, for teachers who have a limited window of time to improve the mathematics performance of students with CAL/WP-MD when early intervention is paramount (Dowker, 2005), WP-RTI provides students with an intervention package that improves a variety of mathematics skills rather than a single aspect of mathematics (i.e., calculations). *Pirate Math*, which is the WP-RTI

program used in the present study, also includes explicit strategy instruction on calculations (note that calculations instruction was not provided in the present study's WP-RTI condition, to permit conclusions about transfer from word-problem instruction to calculations). Therefore, it appears that *Pirate Math* provide teachers with effective step-by-step methods for (a) teaching students how to set-up and solve word problems, (b) embedding calculation strategy instruction practice within word problems, and (c) explicit and implicit practice on pre-algebraic reasoning skills such as understanding operations, setting up equations, and solving for unknowns within equations.

Before automatically discounting CAL-RTI in favor on WP-RTI, however, we emphasize that all small-group tutoring students demonstrated low performance on CAL *and* WP at pretest. While many students with MD experience difficulty in both areas, some students with MD experience difficulty with CAL without WP difficulty or vice versa (Fuchs, Fuchs, Stuebing, et al., 2008; Hanich, Jordan, Kaplan, & Dick, 2001). For students with CAL-MD only, teachers may want to focus on remediation of calculation deficits with CAL-RTI especially if students experience difficulty with double-digit calculations requiring regrouping. With CAL-RTI, however, students did not demonstrate significant improvement on pre-algebraic reasoning. Teachers must provide pre-algebraic preparation in some manner, and results show that such preparation can be effectively provided within the WP-RTI framework using *Pirate Math*.

To effectively implement either CAL-RTI or WP-RTI, school administrators must provide appropriate training for teachers or paraprofessionals. Implementation at both tiers must be conducted with high levels of fidelity to ensure students are receiving appropriate and effective instruction. Additionally, the school schedule must be designed to allocate 45 min of whole-class instruction twice each week and 30 min of small-group tutoring three times each week. This is a large commitment in terms of time and personnel, but if the mathematics performance of the majority of students improves (as in the present study) and the mathematics performance of students most at-risk for MD significantly improves, then the commitment is necessary.

## Acknowledgments

## 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. In: Sutherland, R.; Rojano, T.; Bell, A.; Lins, RC., editors. Perspectives on school algebra. Dordrecht, Netherlands: Kluwer; 2001. p. 249-260.

Baroody AJ, Bajwa NP, Eiland M. Why can't Johnny remember the basic facts? Developmental Disabilities Research Reviews. 2009; 15:69–79.10.1002/ddrr.45 [PubMed: 19213010]

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

Bryant DP, Bryant BR, Gersten R, Scammacca N, Chavez MM. Mathematics intervention for first- and second-grade students with mathematics difficulties: The effects of tier 2 intervention delivered as booster lessons. Remedial and Special Education. 2008; 29:20–32.10.1177/0741932507309712

Bryant DP, Bryant BR, Gersten RM, Scammacca NN, Funk C, Winter A, Pool C. The effects of tier 2 intervention on the mathematics performance of first-grade students who are at risk for mathematics difficulties. Learning Disability Quarterly. 2008; 31:47–63.10.2307/20528817

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

Chong SL, Siegel LS. Stability of computational deficits in math learning disability from second through fifth grades. Developmental Neuropsychology. 2008; 33:300–317.10.1080/87565640801982387 [PubMed: 18473201]

Clement LL, Bernhard JZ. A problem-solving alternative to using key words. Mathematics Teaching in the Middle School. 2005; 10:360–365.

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; 20:405–438.10.1016/0010-0285(88)90011-4

Dowker A. Early identification and intervention for students with mathematics difficulties. Journal of Learning Disabilities. 2005; 38:324–332.10.1177/00222194050380040801 [PubMed: 16122064]

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; 100:829–850.10.1037/a0012657 [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. Mathematical problem-solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. Journal of Learning Disabilities. 2002; 35:563–573.10.1177/00222194020350060701 [PubMed: 15493253]

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; 100:491–509.10.1037/0022-0663.100.3.491 [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. The Elementary School Journal. 1997; 97:223–249.10.1086/461863

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, 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, Schatschneider C, Hamlett CL, Changas P. Effects of first-grade number knowledge tutoring with contrasting forms of practice. Journal of Educational Psychology. 2013; 105:58–77.10.1037/a0030127 [PubMed: 24065865]

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

Fuchs LS, Powell SR, Cirino PT, Schumacher RF, Marrin S, Hamlett CL, Changas PC. Does calculation or word-problem instruction provide a stronger route to pre-algebraic knowledge? Journal of Educational Psychology. in press.

Fuchs LS, Powell SR, Hamlett CL, Fuchs D, Cirino PT, Fletcher JM. Remediating computational deficits at third grade: A randomized field trial. Journal of Research on Educational Effectiveness. 2008; 1:2–32.10.1080/19345740701692449 [PubMed: 21709759]

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; 101:561–576.10.1037/a0014701 [PubMed: 19865600]

Fuchs, LS.; Seethaler, PM.; Powell, SR. Test of Pre-Algebraic Reasoning. L. S. Fuchs; 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. The 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; 23:148–165.10.2307/749498

Geary, DC.; Boykin, AW.; Embretson, S.; Reyna, V.; Siegler, R.; Berch, DB.; Graban, J. Reports of the task groups and subcommittees. Washington, D.C: U.S. Department of Education; 2008. Report of the task group on learning processes. National Mathematics Advisory Panel. Retrieved from http://www2.ed.gov/about/bdscomm/list/mathpanel/report/learning-processes.pdf

Gersten, R.; Beckmann, S.; Clarke, B.; Foegen, A.; March, L.; Star, JR.; Witzel, B. Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools. Washington, D.C: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education; 2009. (NCEE 2009–4060)Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

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

Groen G, Resnick LB. Can preschool children invent addition algorithms? Journal of Educational Psychology. 1977; 69:645–652.10.1037/0022-0663.69.6.645

Hanich LB, Jordan NC, Kaplan D, Dick J. Performance across different areas of mathematics cognition in children with learning difficulties. Journal of Educational Psychology. 2001; 93:615–626.10.1037//0022-0663.93.3.615

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. Solving math word problems: Teaching students with learning disabilities using schema-based instruction. Austin, TX: Pro-Ed; 2007.

Jitendra AK, Hoff K, Beck MM. Teaching middle school students with learning disabilities to solve word problems using a schema-based approach. Remedial and Special Education. 1999; 20:50–64.10.1177/074193259902000108

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, NY: 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; 92:109–129.10.1037/0033-295X.92.1.109 [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; 27:803–812.10.3758/BF03198533 [PubMed: 10540809]

Lembke ES, Garman C, Deno SL, Stecker PM. One elementary school's implementation of response to intervention (RTI). Reading and Writing Quarterly. 2010; 26:361–373.10.1080/10573569.2010.500266

Linchevski, L. Operating on the unknowns: What does it really mean?. In: Heuvel-Panhuizen, Mvd, editor. 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

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.

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]

Mellard DF, Stern A, Woods K. RTI school-based practices and evidence-based models. Focus on Exceptional Children. 2011; 43(6):1–15.

Murnane RJ, Willett JB, Braatz MJ, Duhaldeborde Y. Do different dimensions of male high school students' skills predict labor market success a decade later? Evidence from the NLSY. Economics of Education Review. 2001; 20:311–320.10.1016/S0272-7757(00)00056-X

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; 9:392–389.10.1207/s1532690xci0904_2

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

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; 39:43–51.10.1108/00400919710164125

Peugh JL. A practical guide to multilevel modeling. Journal of School Psychology. 2010; 48:85–112.10.1016/j.jsp.2009.09.002 [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. The Elementary School Journal. 2012; 112:627–648.10.1086/665009 [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]

Powell SR, Fuchs LS. Does early algebraic reasoning differ as a function of students' difficulty with calculations versus word problems? 2014 Manuscript submitted for publication.

Psychological Corporation. Wechsler Abbreviated Scale of Intelligence (WASI) manual. San Antonio, TX: Author; 1999.

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

Riley MS, Greeno JG. Developmental analysis of understanding language about quantities and of solving problems. Cognition and Instruction. 1988; 5:49–101.

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

SAS Institute. SAS 9.3. Cary, NC: Author; 2002–2011.

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.

Singer, JD.; Willett, JB. Applied longitudinal data analysis: Modeling change and event occurrence. Oxford, England: Oxford University Press; 2003.

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

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

Tournaki N, Bae YS, Kerekes J. Rekenrek: A manipulative used to teach addition and subtraction to students with learning disabilities. Learning Disabilities: A Contemporary Journal. 2008; 6(2):41–59.

U.S. Department of Education. Digest of Education Statistics. Washington, D.C: Author; 2008. Retrieved from http://nces.ed.gov/pubs2009/2009020.pdf

Vaughn S, Fuchs LS. Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. Learning Disabilities Research and Practice. 2003; 18:137–146.10.1111/1540-5826.00070

Wei X, Lenz KB, Blackorby J. Math growth trajectories of students with disabilities: Disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. Remedial and Special Education. 2013; 34:154–165.10.1177/0741932512448253

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

**Table 1**

Student Demographics by Tutoring Condition

| Variable | CAL (n = 96) | | WP (n = 91) | | Control (n = 78) | |
|---|---|---|---|---|---|---|
| | **n** | **(%)** | **n** | **(%)** | **n** | **(%)** |
| Males | 47 | (49.0) | 54 | (59.3) | 31 | (39.7) |
| Race | | | | | | |
| African American | 51 | (53.1) | 51 | (56.0) | 41 | (52.6) |
| White | 10 | (10.4) | 14 | (15.4) | 16 | (20.5) |
| Hispanic | 30 | (31.3) | 20 | (22.0) | 14 | (17.9) |
| Other | 5 | (5.2) | 6 | (6.6) | 7 | (9.0) |
| RFL | 90 | (93.8) | 79 | (86.6) | 69 | (88.5) |
| School-identified disability | 8 | (8.3) | 8 | (8.8) | 11 | (14.1) |
| ESL | 18 | (18.8) | 15 | (16.5) | 14 | (17.9) |
| Retained | 7 | (7.3) | 4 | (4.4) | 6 | (11.5) |

*Note.* RFL = Reduced and/or free lunch; ESL = English as a Second Language.

**Table 2**

Performance Data by Treatment Condition

| Variable | CAL Pre M | (SD) | CAL Post M | (SD) | WP Pre M | (SD) | WP Post M | (SD) | Control Pre M | (SD) | Control Post M | (SD) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Screening** | | | | | | | | | | | | |
| Sums to 12 | 3.82 | (1.64) | NA | | 4.04 | (1.61) | NA | | 4.15 | (1.53) | NA | |
| Story Problems | 2.89 | (1.60) | NA | | 3.15 | (1.51) | NA | | 3.12 | (1.51) | NA | |
| WASI IQ | 85.54 | (7.29) | NA | | 87.55 | (9.01) | NA | | 86.65 | (8.72) | NA | |
| **Calculation Outcomes** | | | | | | | | | | | | |
| *Proximal* | | | | | | | | | | | | |
| Sums to 18 | 4.15 | (2.93) | 12.09 | (4.53) | 3.70 | (2.66) | 8.21 | (4.37) | 4.26 | (2.97) | 6.19 | (3.71) |
| Minuends to 12 | 2.98 | (2.18) | 8.53 | (4.64) | 3.04 | (1.90) | 6.86 | (3.21) | 2.95 | (2.02) | 4.65 | (3.12) |
| Minuends to 18 | 2.28 | (2.31) | 8.07 | (3.82) | 2.07 | (2.03) | 4.87 | (3.17) | 2.05 | (2.12) | 3.13 | (2.39) |
| Two-Digit Addition | 1.90 | (2.45) | 14.92 | (5.10) | 1.57 | (2.09) | 8.50 | (4.91) | 1.56 | (1.72) | 6.72 | (5.27) |
| Two-Digit Subtraction | 0.98 | (1.64) | 7.16 | (4.80) | 1.04 | (1.57) | 4.79 | (3.12) | 0.86 | (1.71) | 3.65 | (3.07) |
| *Distal* | | | | | | | | | | | | |
| WRAT Arithmetic[a] | 87.28 | (12.61) | 101.39 | (9.58) | 85.42 | (11.20) | 95.54 | (9.96) | 84.67 | (11.91) | 90.71 | (10.94) |
| KM Addition[a] | NA | | 106.63 | (12.41) | NA | | 98.02 | (10.69) | NA | | 94.10 | (12.45) |
| KM Subtraction[a] | NA | | 97.37 | (15.07) | NA | | 91.98 | (12.42) | NA | | 89.68 | (14.33) |
| **Word-Problem Outcomes** | | | | | | | | | | | | |
| *Proximal* | | | | | | | | | | | | |
| SGSP | 4.92 | (2.88) | 7.91 | (3.89) | 5.07 | (2.24) | 16.05 | (8.11) | 4.30 | (2.18) | 8.59 | (4.39) |
| *Distal* | | | | | | | | | | | | |
| KM Problem Solving[a] | 97.81 | (8.73) | 98.62 | (9.68) | 97.20 | (8.00) | 99.84 | (8.38) | 97.05 | (8.54) | 98.53 | (9.23) |
| ITBS[b] | NA | | 25.75 | (13.58) | NA | | 24.69 | (12.74) | NA | | 24.10 | (12.14) |
| **Pre-Algebraic Reasoning** | | | | | | | | | | | | |
| *Proximal* | | | | | | | | | | | | |
| Find X | NA | | 3.39 | (2.90) | NA | | 6.20 | (1.93) | NA | | 3.13 | (2.83) |

|  | Condition | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  | CAL | | | | WP | | | | Control | | | |
|  | Pre | | Post | | Pre | | Post | | Pre | | Post | |
| Variable | M | (SD) | M | (SD) | M | (SD) | M | (SD) | M | (SD) | M | (SD) |
| Number Sentences | NA | | 0.95 | (1.22) | NA | | 3.11 | (2.27) | NA | | 0.69 | (1.14) |
| Distal | | | | | | | | | | | | |
| DA | NA | | 5.43 | (3.83) | NA | | 5.87 | (3.99) | NA | | 4.44 | (3.84) |
| TPAR | NA | | 8.28 | (3.66) | NA | | 9.77 | (3.81) | NA | | 7.75 | (4.03) |

Note.

[a]Standardized score;

[b]Normal Curve Equivalent.

WASI = Wechsler Abbreviated Scale of Intelligence; WRAT = Wide Range Achievement Test; KM = KeyMath; SGSP = Second-Grade Story Problems; ITBS = Iowa Test of Basic Skills; DA = Dynamic Assessment; TPAR = Test of Pre-Algebraic Reasoning.

**Table 3**

Model Characteristics for Proximal Calculation Outcomes

| Parameter | Single-Digit Proximal | | | | | Double-Digit Proximal | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | (S.E.) | df | t/z/F | 95% CI | Parameter Estimate | (S.E.) | df | t/z/F | 95% CI |
| **Unconditional Model** | | | | | | | | | | |
| Fixed: Intercept ($\gamma_{000}$) | 6.97 | (0.31) | 103 | 22.32[c] | 6.35 to 7.59 | 7.67 | (0.40) | 93.3 | 19.25 | 6.88 to 8.47 |
| Random: Teacher ($\sigma^2_{u0}$) | 5.91 | (1.27) | | 4.65[c] | 4.04 to 9.47 | 9.79 | (2.22) | | 4.41[c] | 6.56 to 16.25 |
| School ($\sigma^2_{v0}$) | | N/A | | | | | N/A | | | |
| Residual ($\sigma^2_e$) | 7.63 | (0.80) | | 9.57[c] | 6.28 to 9.47 | 11.98 | (1.30) | | 9.25[c] | 9.80 to 14.99 |
| Model Fit: | Dev = 1392.2; Param = 2; BIC = 1401.4 | | | | | Dev = 1503.7; Param = 2; BIC = 1512.9 | | | | |
| **Full/Final Model** | | | | | | | | | | |
| Fixed: Intercept ($\gamma_{000}$) | 2.87 | (0.63) | 134 | 4.58[c] | 1.63 to 4.11 | 3.14 | (0.93) | 110 | 3.36[b] | 1.29 to 4.99 |
| Pretest | 0.53 | (0.10) | 234 | 5.25[c] | 0.33 to 0.73 | 1.69 | (0.57) | 246 | 2.98[b] | 0.57 to 2.80 |
| Treatment (CAL) | 4.45 | (0.58) | 85.1 | 7.69[c] | 3.31 to 5.60 | 5.57 | (0.71) | 64 | 7.81[c] | 4.15 to 7.00 |
| Treatment (WP) | 1.86 | (0.58) | 91 | 3.17[b] | 0.70 to 3.02 | 1.24 | (0.72) | 66.4 | 1.72 | −0.20 to 2.68 |
| Cohort (1) | −0.38 | (0.62) | 110 | −0.62 | −1.62 to 0.85 | −0.68 | (0.80) | 35.5 | −0.85 | −2.31 to 0.95 |
| Cohort (2) | 0.73 | (0.65) | 136 | 1.13 | −0.55 to 2.02 | 0.39 | (0.87) | 42.8 | 0.45 | −1.36 to 2.13 |
| Cohort (3) | 1.09 | (0.62) | 133 | 1.76 | −0.14 to 2.31 | 1.63 | (0.79) | 26.6 | 2.06[a] | 0.01 to 3.26 |
| Random: Teacher ($\sigma^2_{u0}$) | 2.01 | (0.70) | | 2.88[b] | 1.12 to 4.57 | 2.07 | (1.32) | | 1.57 | 0.80 to 12.70 |
| School ($\sigma^2_{v0}$) | | N/A | | | | 0.16 | (0.63) | | 0.25 | 0.01 to 5.80 |
| Residual ($\sigma^2_e$) | 6.84 | (0.72) | | 9.50[c] | 5.62 to 8.50 | 12.10 | (1.33) | | 9.11[c] | 0.44 to 0.53 |
| Model Fit: | Dev = 1279.4; Param = 2; BIC = 1288.6 | | | | | Dev = 1378.2; Param = 3; BIC = 1387.9 | | | | |

Note.

[a] $p < .05$;

[b] $p < .01$;

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Pretest refers single-digit and double-digit computation, respectively. CAL is calculation intervention. WP is word-problem intervention. For the unconditional model, *t* values are reported. Other variables considered but not needed in final model (sex, lunch status, ethnicity, second language status, special education status, age) are not shown above. 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 (reported) are all 1; denominators (not included) are all 1; 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 and for Cohort 4. For model fit, Dev = deviance or −2log likelihood; Param = (random) parameters; BIC = Bayesian Information Criteria. For Dev. and BIC, lower values are better fit. Intraclass correlations (ICCs) are reported in text.

**Table 4**

Model Characteristics for Distal Calculation Outcome

| Parameter | Distal (Standardized) Calculation | | | | |
|---|---|---|---|---|---|
| | Parameter Estimate | (S.E.) | df | t/z/F | 95% CI |
| **Unconditional Model** | | | | | |
| Fixed: Intercept ($\gamma_{000}$) | −0.00 | (0.07) | 21.9 | −0.01 | −0.14 to 0.15 |
| Random: Teacher ($\sigma^2_{u0}$) | 0.14 | (0.06) | | 2.38[b] | 0.07 to 0.38 |
| School ($\sigma^2_{v0}$) | 0.02 | (0.03) | | 0.71 | 0.00 to 24.13 |
| Residual ($\sigma^2_e$) | 0.53 | (0.06) | | 9.55[c] | 0.43 to 0.65 |
| **Full/Final Model** | | | | | |
| Model Fit: | Dev = 636.2; Param = 3; BIC = 645.9 | | | | |
| Fixed: Intercept ($\gamma_{000}$) | 0.57 | (0.92) | 244 | 0.63 | −1.23 to 2.38 |
| Pretest | 0.03 | (0.00) | 245 | 8.66[c] | 0.02 to 0.04 |
| Age | −0.44 | (0.10) | 243 | −4.47[c] | −0.63 to −0.25 |
| Spec. Ed. (Receive) | −0.31 | (0.11) | 244 | −2.67[b] | −0.54 to −0.08 |
| Treatment (CAL) | 0.73 | (0.09) | 67.8 | 8.07[c] | 0.55 to 0.91 |
| Treatment (WP) | 0.26 | (0.09) | 74.9 | 2.90[b] | 0.08 to 0.44 |
| Cohort (1) | −0.11 | (0.13) | 36.2 | −0.89 | −0.37 to 0.15 |
| Cohort (2) | −0.04 | (0.14) | 45.2 | −0.27 | −0.31 to 0.24 |
| Cohort (3) | 0.01 | (0.13) | 33.8 | 0.06 | −0.26 to 0.28 |
| Random: Teacher ($\sigma^2_{u0}$) | 0.00 | (0.02) | | 0.31 | 0.00 to −0.00 |
| School ($\sigma^2_{v0}$) | 0.03 | (0.02) | | 1.64 | 0.01 to 0.15 |
| Residual ($\sigma^2_e$) | 0.27 | (0.03) | | 9.70[c] | 0.22 to 0.33 |
| Model Fit: | Dev = 436.0; Param = 3; BIC = 445.6 | | | | |

*Note.* See note for Table 3. The pretest was WRAT Arithmetic. The code 0 is for the treatment control group, for no special education, and for Cohort 4.

**Table 5**

Model Characteristics for Proximal and Distal Word-Problems Outcomes

| Parameter | Proximal | | | | | Distal | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Estimate | (S.E.) | Df | t/z/F | 95% CI | Parameter Estimate | (S.E.) | Df | t/z/F | 95% CI |
| **Unconditional Model** | | | | | | | | | | |
| Fixed: Intercept ($\gamma_{000}$) | 10.98 | (0.59) | 87.3 | 18.72c | 9.82 to 12.15 | −0.01 | (0.06) | 24.2 | −0.12 | −0.12 to 0.11 |
| Random: Teacher ($\sigma^2_{u0}$) | 20.54 | (5.00) | N/A | 4.11c | 13.42 to 35.34 | | | N/A | | |
| School ($\sigma^2_{v0}$) | | | | | | 0.03 | (0.02) | | 1.51 | 0.01 to 0.19 |
| Residual ($\sigma^2_e$) | 27.56 | (3.01) | | 9.17c | 22.50 to 34.55 | 0.40 | (0.04) | | 11.07c | 0.34 to 0.48 |
| Model Fit: | Dev = 1729.0; Param = 2; BIC = 1738.1 | | | | | Dev = 520.9; Param = 2; BIC = 527.3 | | | | |
| **Full/Final Model** | | | | | | | | | | |
| Fixed: Intercept ($\gamma_{000}$) | 6.46 | (1.28) | 65.3 | 5.06c | 3.91 to 9.01 | 2.77 | (0.69) | 246 | 4.01c | 1.41 to 4.13 |
| Pretest | 0.73 | (0.12) | 236 | 6.12c | 0.50 to 0.97 | 0.53 | (0.06) | 245 | 8.36c | 0.41 to 0.66 |
| Spec. Ed. (Receive) | −2.20 | (1.05) | 241 | −2.09a | −4.27 to −0.13 | | | N/A | | |
| Age | | | N/A | | | −0.38 | (0.09) | 246 | −4.19c | −0.55 to −0.20 |
| Treatment (CAL) | −2.38 | (1.55) | 71.7 | −1.54 | −5.47 to 0.71 | −0.05 | (0.09) | 246 | −0.50 | −0.22 to 0.13 |
| Treatment (WP) | 6.82 | (1.72) | 88.9 | 3.96c | 3.40 to 10.23 | −0.01 | (0.09) | 246 | −0.16 | −0.20 to 0.17 |
| Cohort (1) | −0.89 | (1.71) | 50.9 | −0.52 | −4.32 to 2.54 | 0.11 | (0.12) | 38.6 | 0.86 | −0.14 to 0.36 |
| Cohort (2) | −1.61 | (1.84) | 61 | −0.88 | −5.29 to 2.06 | 0.23 | (0.13) | 46.4 | 1.76 | −0.03 to 0.50 |
| Cohort (3) | −1.31 | (1.83) | 61.9 | −0.72 | −4.98 to 2.35 | 0.06 | (0.13) | 35 | 0.51 | −0.19 to 0.32 |
| Random: Teacher ($\sigma^2_{u0}$) | 0.81 | (2.02) | | 0.40 | 0.10 to 2.28 | | | N/A | | |
| School ($\sigma^2_{v0}$) | 1.18 | (1.19) | | 0.99 | 0.32 to 49.06 | 0.02 | (0.01) | | 1.43 | 0.01 to 0.16 |
| Residual ($\sigma^2_e$) | 22.73 | (2.57) | | 8.86c | 18.43 to 28.7 | 0.30 | (0.03) | | 10.77c | 0.25 to 0.36 |
| Model Fit: | Dev = 1516.3; Param = 3; BIC = 1526.0 | | | | | Dev = 440.7; Param = 2; BIC = 447.2 | | | | |

*Note.* See note for Table 3. Pretest refers to Second-Grade Story Problems (Proximal) or Word Problems Standardized Composite (Distal). Estimates for the significant treatment by cohort interaction in left-hand model not shown, but were included in model that produced the above numbers; results described in text. The code 0 is for the treatment control group, for no special education, and for Cohort 4.

**Table 6**

Model Characteristics for Proximal and Distal Pre-Algebra Outcomes

| Parameter | Proximal | | | | | Distal | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Estimate | (S.E.) | df | t/z/F | 95% CI | Estimate | (S.E.) | df | t/z/F | 95% CI |
| **Unconditional Model** | | | | | | | | | | |
| Fixed: Intercept ($\gamma_{000}$) | 0.01 | (0.06) | 92.3 | 0.16 | −0.10 to 0.12 | −0.04 | (0.06) | 20.6 | −0.67 | −0.18 to 0.09 |
| Random: Teacher ($\sigma^2_{u0}$) | 0.17 | (0.04) | | 3.88[c] | 0.11 to 0.30 | 0.09 | (0.05) | | 1.61 | 0.03 to 0.49 |
| School ($\sigma^2_{v0}$) | | | N/A | | | 0.01 | (0.02) | | 0.21 | 0.00 to 7.30 |
| Residual ($\sigma^2_e$) | 0.29 | (0.03) | | 9.30[c] | 0.24 to 0.37 | 0.67 | (0.07) | | 9.71[c] | 0.55 to 0.82 |
| Model Fit: | Dev = 511.3; Param = 2; BIC = 520.4 | | | | | Dev = 673.9; Param = 3; BIC = 683.6 | | | | |
| **Full/Final Model** | | | | | | | | | | |
| Fixed: Intercept ($\gamma_{000}$) | −0.30 | (0.09) | 98.8 | −3.40[c] | −0.47 to −0.12 | −1.10 | (0.41) | 239 | −2.72[b] | −1.90 to −0.30 |
| Word-Problem Pretest | 0.24 | (0.06) | 247 | 3.99[c] | 0.12 to 0.36 | 0.48 | (0.10) | 247 | 5.04[c] | 0.29 to 0.67 |
| WRAT-3 Pretest | | | N/A | | | 0.01 | (0.00) | 246 | 2.50[a] | 0.00 to 0.02 |
| Spec. Ed. (Receive) | −0.31 | (0.11) | 247 | −2.78[b] | −0.53 to −0.09 | −0.46 | (0.16) | 247 | −2.81[b] | −0.79 to −0.14 |
| Treatment (CAL) | 0.02 | (0.09) | 80.5 | 0.25 | −0.16 to 0.20 | 0.10 | (0.13) | 69.5 | 0.73 | −0.17 to 0.36 |
| Treatment (WP) | 0.83 | (0.09) | 86.1 | 8.98[c] | 0.65 to 1.02 | 0.32 | (0.13) | 72.4 | 2.47[a] | 0.06 to 0.59 |
| Cohort (1) | 0.06 | (0.10) | 111 | 0.56 | −0.14 to 0.26 | −0.23 | (0.16) | 44.2 | −1.47 | −0.55 to 0.09 |
| Cohort (2) | −0.06 | (0.11) | 123 | −0.53 | −0.27 to 0.16 | 0.14 | (0.17) | 49.6 | 0.85 | −0.20 to 0.48 |
| Cohort (3) | 0.08 | (0.10) | 126 | 0.79 | −0.12 to 0.27 | 0.05 | (0.16) | 37.6 | 0.32 | −0.27 to 0.37 |
| Random: Teacher ($\sigma^2_{u0}$) | 0.02 | (0.02) | | 1.11 | 0.01 to 0.40 | 0.03 | (0.04) | | 0.79 | 0.01 to 8.30 |
| School ($\sigma^2_{v0}$) | | | N/A | | | 0.02 | (0.02) | | 0.77 | 0.00 to 5.36 |
| Residual ($\sigma^2_e$) | 0.26 | (0.03) | | 9.61[c] | 0.21 to 0.32 | 0.53 | (0.06) | | 9.47[c] | 0.43 to 0.66 |
| Model Fit: | Dev = 412.6; Param = 2; BIC = 421.7 | | | | | Dev = 600.8; Param = 3; BIC = 610.5 | | | | |

*Note.* See note for Table 3. The code 0 is for the treatment control group, for no free/reduced lunch, for no special education, and for Cohort 4.