



Published in final edited form as:

J Educ Psychol. 2009 August 1; 101(3): 561–576. doi:10.1037/a0014701.

Remediating Number Combination and Word Problem Deficits Among Students With Mathematics Difficulties: A Randomized Control Trial

Lynn S. Fuchs¹, Sarah R. Powell¹, Pamela M. Seethaler¹, Paul T. Cirino², Jack M. Fletcher², Douglas Fuchs¹, Carol L. Hamlett¹, and Rebecca O. Zumeta¹

¹ Department of Special Education, Vanderbilt University

² Department of Psychology, University of Houston

Abstract

The purposes of this study were to assess the efficacy of remedial tutoring for 3rd graders with mathematics difficulty, to investigate whether tutoring is differentially efficacious depending on students' math difficulty status (mathematics difficulty alone vs. mathematics plus reading difficulty), to explore transfer from number combination (NC) remediation, and to examine the transportability of the tutoring protocols. At 2 sites, 133 students were stratified on mathematics difficulty status and site and then randomly assigned to 3 conditions: control (no tutoring), tutoring on automatic retrieval of NCs (i.e., Math Flash), or tutoring on word problems with attention to the foundational skills of NCs, procedural calculations, and algebra (i.e., Pirate Math). Tutoring occurred for 16 weeks, 3 sessions per week and 20–30 min per session. Math Flash enhanced fluency with NCs with transfer to procedural computation but without transfer to algebra or word problems. Pirate Math enhanced word problem skill as well as fluency with NCs, procedural computation, and algebra. Tutoring was not differentially efficacious as a function of students' mathematics difficulty status. The tutoring protocols proved transportable across sites.

Keywords

mathematics disability; validated mathematics tutoring; word problems; number combinations

Mathematics competence accounts for variance in employment, income, and work productivity even after intelligence and reading have been explained (Rivera-Batiz, 1992). So it is unfortunate that mathematics disability is widespread, affecting 5%–9% of the school-age population (e.g., Badian, 1983; Gross-Tsur, Manor, & Shalev, 1996). Together, the lifelong challenges associated with mathematics disability and the high prevalence of the disorder make mathematics disability a critical public health problem. For this reason, it is essential to prevent mathematics difficulties.

Research shows that early prevention activities can substantially improve math performance (e.g., Clements & Sarama, 2007; Fuchs, Fuchs, Yazdian, & Powell, 2002; Griffin, Case, & Siegler, 1994). Yet there are no interventions that are effective for all students. In Fuchs et al. (2005), for example, a first-grade prevention program reduced the prevalence of mathematics disability at the end of first grade, with effects maintaining 1 year after tutoring ended

(Compton, Fuchs, & Fuchs, 2007). Even so, a subset of tutored students, approximately 3%–6% of the school population, continued to manifest severe mathematics deficits. Because we cannot expect prevention activities to be universally effective, the need for intensive remedial intervention persists even when strong prevention services are available.

In the present study, we focused on the remediation of mathematics delays at third grade, when serious mathematics deficits are clearly established and identification of mathematics disability begins (Fletcher, Lyon, Fuchs, & Barnes, 2007). With a randomized control trial, we assessed the efficacy of two tutoring protocols: one for remediating number combination deficits and the other for remediating word problem deficits.¹ We examined efficacy as a function of the nature of mathematics difficulty: whether it occurs alone or in combination with reading problems. In this introduction, for each aspect of mathematical cognition, we summarize previous remediation work and explain the theoretical underpinnings of the questions we posed and the approaches to remediation we employed. Then we review the purpose of the present study.

Number Combinations

Number combinations (NCs) are simple arithmetic problems (e.g., $5 + 7 = 12$, $9 - 5 = 4$) that can be solved via counting or decomposition strategies or committed to long-term memory for automatic retrieval. Consensus exists that NC skill is essential (Kilpatrick, Swafford, & Findell, 2001), and research shows that fluency with NCs is a significant path to procedural computation and word problem performance (Fuchs, Fuchs, Compton et al., 2006). To answer addition NCs, typical children gradually develop procedural efficiency with counting. First they count two sets (e.g., $2 + 3$) in their entirety (i.e., 1, 2, 3, 4, 5); then they count from the first addend (i.e., 2, 3, 4, 5); and eventually they count from the larger addend (i.e., 3, 4, 5). As conceptual knowledge about number becomes more sophisticated, individuals also develop decomposition strategies for deriving answers (e.g., $[2 + 2 = 4] + 1 = 5$). As increasingly efficient counting and decomposition strategies help individuals consistently and quickly pair problems with correct answers in working memory, associations become established in long-term memory, and individuals gradually favor memory-based retrieval of answers (Ashcraft & Stazyk, 1981; Geary, Widaman, Little, & Cormier, 1987; Goldman, Pellegrino, & Mertz, 1988; Groen & Parkman, 1972; Siegler, 1987).

Students with mathematics disability manifest greater difficulty with counting (Geary, Bow-Thomas, & Yao, 1992; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007); persist with immature backup strategies (Geary et al., 2007); and fail to make the shift to memory-based retrieval of answers (Fleischner, Garnett, & Shepherd, 1982; Geary et al., 1987; Goldman et al., 1988). When children with mathematics disability do retrieve answers from memory, they commit more errors and manifest unsystematic retrieval speeds than younger, typically developing counterparts (Geary, Brown, & Samaranayake, 1991; Gross-Tsur et al., 1996; Ostad, 1997). Some (e.g., Fleischner et al., 1982; Geary et al., 1987; Goldman et al., 1988) consider NCs to be a signature deficit of students with mathematics disability, and difficulty with automatic retrieval of NCs is one of the most consistent findings in the mathematics disability literature (e.g., Cirino, Ewing-Cobbs, Barnes, Fuchs, & Fletcher, 2007; Geary et al., 2007; Jordan, Hanich, & Kaplan, 2003).

Conventionally, NCs are incorporated into the curriculum at kindergarten through second grade, although many general educators do not devote explicit attention toward developing

¹We refer to NC tutoring as Math Flash because NCs “flash” during computerized practice. We refer to WP tutoring as Pirate Math because the materials incorporate a pirate theme. For information on how to obtain manuals with the tutoring scripts, contact flora.murry@vanderbilt.edu

strategies for solving NCs or promoting fluency with NCs (Miller & Hudson, 2007). Even so, typically developing students are fluent with NCs by the beginning of third grade (Cirino et al., 2007). When students still manifest deficiencies at the beginning of third grade, a pressing need exists for remediation.

The research literature on remediation of NC deficits is limited. Okolo (1992) and Christensen and Gerber (1990) contrasted computerized practice with NCs for students with learning disabilities in a game versus a drill format. Okolo found no significant differences between groups, both of which improved, whereas Christensen and Gerber found that students were disadvantaged by the game format, perhaps because of its distracting nature. Neither of these two studies, however, incorporated a control group to assess whether computerized practice effected better outcomes than business as usual. In a third study, which was also conducted without a control group, Tournaki (2003) contrasted paper-pencil drill and practice with instruction designed to teach students with learning disabilities to count NC answers strategically. Results showed an advantage for strategic counting. This finding was difficult to interpret because paper-pencil practice provided feedback on a delayed schedule, without mixing known with unknown NCs and without systematic review. By contrast, strategy instruction incorporated immediate corrective feedback, systematic review, and reteaching whenever errors occurred.

These prior studies therefore fail to provide the basis for determining whether NC remediation leads to better progress than would be expected with business-as-usual schooling. Also, participants in this prior work had school-identified learning disabilities, making it difficult to determine whether effects apply specifically to students with mathematics difficulty. The present study adds to the literature methodologically by incorporating random assignment, including a control condition, and screening participants to confirm mathematics difficulty. Moreover, to extend the literature substantively, we were interested in whether the efficacy of NC remediation differed as a function of whether mathematics difficulty occurs alone or in combination with reading difficulty—a scheme that has been proposed for subtyping mathematics disability. As Geary (1993) hypothesized, because a key deficit associated with reading difficulty is phonological processing (Bruck, 1992) and because phonological processing deficits are linked to difficulty with automatic retrieval of NCs (Fuchs et al., 2005), students with concurrent difficulty in mathematics and reading should experience greater difficulty with NCs compared with students who experience difficulty with mathematics alone.

Research suggests that compared with students with concurrent difficulty, those with mathematics difficulty alone use more efficient counting procedures to solve NCs (Geary, Hamson, & Hoard, 2000; Jordan & Hanich, 2000) with faster retrieval times (Andersson & Lyxell, 2007; Hanich et al., 2001; Jordan & Montani, 1997) but comparable accuracy (Cirino et al., 2007). The literature is not, however, consistent (e.g., Micallef & Prior, 2004; Reikeras, 2006), and most studies addressing these questions have employed a cross-sectional causal-comparative design. An alternative approach for studying the same issue is experimental, in which students with these subtypes are randomly assigned to treatment or control conditions with the goal of determining whether the subtypes respond differentially to intervention. This design offers the basis for stronger, causal inferences about the tenability of the subtyping scheme.

In the present study, we adopted this methodological approach. The NC remediation relied primarily on counting strategies and practice, although we also addressed adding and subtracting concepts, the commutative property of addition, and the concepts of 1 and 0. We taught students the min strategy for adding (start with the larger addend; count up to the other addend; the answer is the last number counted) and the missing-addend strategy for subtracting

(start with the minus number; count up to the starting number; the answer is the number of counts or fingers up). We provided practice to develop fluency with these counting strategies and help instantiate NCs in long-term memory. Using this remediation protocol, which we had not previously tested, we hypothesized that students with mathematics difficulty alone would prove more responsive to this NC remediation than those with concurrent reading difficulty on the basis of (a) evidence indicating that third graders with concurrent math and reading difficulty are as accurate with NCs but manifest slower retrieval times on small sums than those with mathematics difficulty alone (Cirino et al., 2007) and (b) Geary's (1993) theoretical framework suggesting that the phonological processing deficits associated with reading disability are linked to difficulty with automatic retrieval of NCs.

Finally, we were also interested in whether NC remediation would transfer to procedural calculations or word problems. Research shows that many children experience difficulty transferring the math competence they develop in school (Foxman, Ruddock, McCallum, & Schagen, 1991, as cited in Boaler, 1993; Larkin, 1989). The issue of transfer from NCs to other aspects of mathematics is theoretically interesting in the field of mathematics disability because NCs are viewed as a signature deficit, representing a bottleneck for students with mathematics disability (Fleischner et al., 1982; Geary et al., 1987; Goldman et al., 1988). The hypothesis is that with a fixed amount of attention, students with NC deficits allocate available resources for deriving answers to these simple problems instead of focus on the more complex mathematics into which NCs are embedded (cf. Ackerman, Anhalt, & Dykman, 1986; Goldman & Pellegrino, 1987). In contrast, in the mathematics education literature, transfer difficulties are interpreted within a theoretical framework that challenges the assumption of vertical transfer, whereby mastery of simple skills facilitates acquisition of more complex skills (Gagne, 1968; Resnick & Resnick, 1992). Findings of the present study should lend support to one of these competing theoretical frameworks and should also provide insight into whether NCs are a signature deficit or simply represent one component among a constellation of difficulties.

Word Problems

In contrast to NCs, which are already set up for solution, word problems (WPs) require students to use text to identify missing information, construct the number sentence, derive the calculation problem for finding the missing information, and finally solve that calculation problem. The need for text to construct the problem model appears to alter the nature of the task. Some lines of research suggest that computation and WPs may represent distinct aspects of mathematical cognition (e.g., Fuchs, Fuchs, Compton, et al., 2006; Fuchs, Fuchs, Stuebing, et al., 2008; Swanson, 2006; Swanson & Beebe-Frankenberger, 2004), so that computation and WP skill need to be considered separately in identifying and remediating students with mathematics disabilities.

The major approach in the research literature for developing WP skill for students with learning difficulties relies on schema theory, which is based on the concept of lateral transfer by which children recognize problems across numerous experiences to abstract generalized problem-solving strategies (Resnick & Resnick, 1992). Some refer to the abstraction of generalized problem-solving strategies as the development of schemas (Brown, Campione, Webber, & McGilly, 1992; Cooper & Sweller, 1987). A *schema* is a category that encompasses similar problems; it is a problem type (Chi, Feltovich, & Glaser, 1981; Quilici & Mayer, 1996). For example, in the Appendix, we show the problem types we addressed in the present study: the Total problem type, the Difference problem type, and the Change problem type (see also explanations in the Method section). Instruction based on schema theory encourages students to develop a schema for each of these problem types. The broader the category for the problem type (i.e., the broader the schema), the greater the probability that students will recognize a novel problem as belonging to a familiar problem type (i.e., schema) for which they know a

solution method. To facilitate schema development, teachers must first teach problem-solution rules. Then teachers must help students develop schemas for the problem types and awareness of those schemas (Cooper & Sweller, 1987). Broadening schemas should affect breadth of learning or transfer (Brown et al., 1992; Glaser, 1983). Research has substantiated the importance of mastering rules for problem solution (e.g., Mawer & Sweller, 1985), but less is known about how to help students develop schemas and awareness of those schemas (e.g., Bransford & Schwartz, 1999; Cooper & Sweller, 1987).

In the past decade, some research programs have relied on explicit instruction based on schema theory to enhance WP skill. Jitendra and colleagues demonstrated acquisition, maintenance, and transfer effects for students with serious mathematics deficits or with risk for mathematics difficulty at eighth grade (Jitendra, DiPipi, & Perron-Jones, 2002), sixth grade (Xin, Jitendra, & Deatline-Buchman, 2005), and third and fourth grades (Jitendra et al., 1998, 2007; Jitendra & Hoff, 1996). In our intervention work on WPs, we have also relied on schema theory. Similar to Jitendra, we teach students to understand the underlying mathematical structure of the problem type, to recognize the basic problem type, and to solve the problem type. In contrast to Jitendra, we incorporate a fourth instructional component by explicitly teaching students to transfer their WP skills. In keeping with Cooper and Sweller (1987), our goal with this fourth instructional component is to help students recognize connections between problems such as those worked during instruction and problems with unexpected features. Unexpected features can include, for example, irrelevant information or novel questions that require an extra step or relevant information presented in charts or graphs or combinations of problem types. We refer to these unexpected features as *transfer features*. In our work, we have addressed these and other transfer features. The addition of explicit instruction on transfer features should lead to more flexible and successful problem solving. We refer to the combination of all four instructional components as *schema-broadening instruction*, or SBI.

In our first randomized controlled study, Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, et al. (2003) isolated the effects of our fourth instructional component (explicitly teaching for transfer) from the first three instructional components (teaching students to understand the underlying mathematical structure of the problem type, to recognize the basic problem type, and to solve the problem type). Working with third graders without mathematics difficulties, we found that SBI (i.e., all four components) strengthened WP performance beyond experimenter-designed instruction on the first three instructional components, including performance on a far-transfer measure that required students to solve taught and untaught problem types in a highly novel and complex context that resembled real-life problem solving. In a series of additional studies on SBI, also conducted in general education (Fuchs, Fuchs, Craddock, et al., 2008; Fuchs, Fuchs, Finelli, Courey, & Hamlett, 2004; Fuchs, Fuchs, Finelli, Courey, Hamlett, et al., 2006; Fuchs, Fuchs, Prentice, Burch, Hamlett, Owen, & Schroeter, 2003; Fuchs, Fuchs, Prentice, Hamlett, et al., 2004), effect sizes favoring SBI were large (0.89–2.14). Random assignment, however, occurred at the classroom level, with limited numbers of students with mathematics difficulty included.

Most recently, Fuchs, Seethaler, et al. (2008) piloted SBI, this time conducted as tutoring rather than whole-class instruction, for third graders whom we identified as having mathematics and reading difficulties. The 35 participants, who scored on average at the 10th percentile in math and reading, were randomly assigned to receive SBI tutoring or to continue in their mathematics program without modification. Problems were less complex than in previous phases of the research program, limited to one-step Total, Difference, and Change problem types as in the present study (see Appendix). Results favored WP performance among the tutored students, but instructional time across the tutored and control students was not controlled—a limitation we address in the present study.

In the present study, our goal was not to extend schema theory or to contrast SBI to alternative approaches for enhancing WP performance. Rather, in the present study, we used SBI as a validated approach with which we could examine a different theoretical issue: whether students respond differentially to WP tutoring as a function of whether they experience mathematics difficulty alone or in combination with reading difficulty. Using text to construct the WP model appears to involve language. A cognitive characteristic associated with WP skill is language (e.g., Fuchs et al., 2005; Fuchs, Fuchs, Compton, et al., 2006; Fuchs, Fuchs, Stuebing, et al., 2008; Swanson, 2006; Swanson & Beebe-Frankenberger, 2004), and the language profiles of students with concurrent difficulty with mathematics and reading are lower than those for students who experience mathematics difficulty alone (Powell, Fuchs, Fuchs, Cirino, & Fletcher, in press). This pattern of performance is evident on simple (e.g., Hanich, Jordan, Kaplan, & Dick, 2001; Jordan & Hanich, 2000; Powell et al., in press) and complex (Fuchs & Fuchs, 2002) WPs. In related experimental work, Fuchs, Fuchs, and Prentice (2004) randomly assigned third-grade classrooms to validated or control WP instruction. Students were retrospectively identified with difficulty in mathematics and reading, with mathematics alone, with reading alone, or without either form of difficulty. After 16 weeks of intervention, students with concurrent difficulty in mathematics and reading were less responsive on problem-solving scores than students without difficulty. In this way, finding differential response to validated instruction for students with mathematics difficulty alone versus those with concomitant reading difficulty suggests the viability of the subtyping scheme, but the study was methodologically limited because of the retrospective assignment to subtypes and limited statistical power, issues addressed in the present study.

Purpose of the Present Study

In the present study, we conducted a randomized field trial in which students were stratified on site and difficulty status and then randomly assigned to a control group or one of the two tutoring conditions. We had four purposes. The first was to assess the efficacy of a tutoring protocol for remediating NC deficits and a tutoring protocol for remediating WP deficits. These tutoring protocols were evaluated against each other and against a no-tutoring control group. By including a control group, we controlled for maturation, historical effects, and business-as-usual schooling. By incorporating two tutoring conditions, we controlled for tutoring time when considering the effects of one protocol against the other. A second and related purpose was to examine the transportability of these tutoring protocols. We accomplished this by conducting the study at two sites, one of which was distal to the developers.

Our second and third purposes were more theoretical. We explored whether tutoring in each aspect of mathematical cognition was differentially efficacious depending on students' difficulty status: mathematics difficulty alone versus mathematics difficulty with concomitant reading difficulty. Findings should advance understanding of Geary's (1993) subtyping scheme, which has helped guide research on mathematics disability over the past decade, but largely via causal-comparative designs. Finally, we investigated the issue of transfer from NCs to procedural calculations and WPs. This should help determine whether NCs represent a signature and bottleneck deficit or instead constitute one aspect of mathematics disability. (Note that although we assessed transfer from NC tutoring to procedural calculation and WP outcomes, we did not examine transfer from WPs to NCs because this latter issue is theoretically less interesting. Also, our design precluded this focus because WP tutoring addressed mathematics deficits foundational to solving WPs, including a counting strategy to derive answers to NCs.)

Method

Participants

The study was conducted at two sites, both large urban school districts. Houston was distal and Nashville was proximal to the developers of the tutoring protocols. Third-grade students ($n = 924$) were screened for inclusion in 63 classrooms in 18 schools. Seven schools and 23 classrooms were in Houston; 11 schools and 40 classrooms were in Nashville. Because tutoring focused on NCs or on WPs, we included students with low performance on a calculations screening measure or a WP screening measure. (Screening occurred in stepwise fashion, so students did not receive every measure.) The criterion applied for low performance on the calculations measure was less than the 26th percentile. The criterion applied to the five-item word problem measure was a score of 0 or 1. (See *Measures* for description of the screening measures.) All 924 students were administered the calculations measure; 302 (33%) scored less than the 26th percentile. We administered the five-item WP screener to 598 students; 170 (28%) scored 0 or 1. Of the 598 students who took the calculations and WP screening measures, 291 (49%) did not meet the inclusion criterion on either measure, 67 (11%) met only the WP criterion, 137 (23%) met only the calculations criterion, and 103 (17%) met both criteria.

The 307 students who met either or both criteria were eligible for further screening on a reading and an abbreviated IQ measure. We excluded students who scored between the 25th and 40th percentiles in reading and students with a T score below 30 on both IQ subtests. Students scoring less than the 26th percentile on the reading measure were classified as having math and reading difficulty (MDRD). Those scoring more than the 39th percentile were classified as math difficulty alone (MD). Two hundred two students took all measures. Of these students, 32 (16%) were excluded because of reading scores between the 25th and 40th percentiles, 2 students were excluded because of low IQ scores, and 1 student was excluded for both reasons. Thus, 165 students were eligible for tutoring. However, 162 students composed the actual assignment sample because 3 students who met all criteria were accidentally not included in the assignment sample.

Blocking on site, type of screening difficulty (WPs, calculations, or both), and difficulty status (MD or MDRD), we randomly assigned students to one of three treatment conditions (NC tutoring, WP tutoring, or control). So, the composition of each treatment group was similar in terms of the three blocking variables. Of the 162 students, 13 (8%) moved after randomization but prior to the onset of tutoring, 7 (4%) moved during the school year, 5 (3%) were excluded by parents or schools prior to the onset of tutoring, and 4 (2%) were withdrawn by parents or schools during the school year, leaving 133 who were evaluated at posttest.

The mean age of the sample ($n = 133$) was 8.94 years ($SD = 0.54$). Sixty-seven students were in Houston; 66 in Nashville. Seventy-three (55%) met criteria for MD; 60 (45%) for MDRD. Fifty-eight (44%) met screening criteria only on the calculation measure; 16 (12%) only on the WP measure; 59 (44%) on both measures. Forty-four students (33%) were assigned to NC tutoring, 42 (32%) to WP tutoring, and 47 (35%) to the control group. The number of minutes of tutoring approximated 18.5 hr (1,099 min; $SD = 148$).

Twenty-one students (16%) were classified as English learners (data were missing for 6 students). Seventy-three (55%) were boys (data were missing for 2 students). Thirty-six (27%) had been retained (data were missing for 6). One hundred four (78%) were eligible for subsidized lunch (data were missing for 2). Twenty-three (17%) were classified as special education (data were missing for 7). Fifty-four (63%) were African American, 31 (23%) Hispanic, 12 (9%) Caucasian, and 3 (2%) Asian (the other 3 students were of other ethnicity or their data were missing).

As expected, given the randomization procedure, treatment groups did not differ by site, difficulty status (MD vs. MDRD), or qualifying criteria (both vs. calculations vs. WPs; all $p > .05$). Treatment groups also did not differ in age, second-language status, gender, free lunch status, special education status, retained status, or ethnicity (all $p > .05$). Table 1 provides demographic and screening data for the sample by treatment condition and by difficulty status.

Classroom Mathematics Program

At both sites, NC instruction was minimal. WP instruction addressed the three problem types taught in tutoring as well as more complex problem types. Problem types were addressed one at a time and focused on underlying concepts and solution strategies. There was no attempt to broaden students' schemas to address transfer. In Nashville, the classroom program was Houghton Mifflin Math (Greenes et al., 2005). A prescribed set of problem-solution rules was taught, with explicit steps for arriving at solutions. Students were provided with guiding questions to help them understand, plan, solve, and reflect on the content of problems. In comparison to WP tutoring, classroom instruction provided more practice in applying problem-solution rules and greater emphasis on computational requirements. Classroom instruction was explicit and relied on worked examples, guided group practice, independent work with checking, and homework. Houston permitted schools to select their own classroom mathematics program but required that instruction be guided by its Horizontal Alignment Planning Guide, which was aligned to the high-stakes test. Instruction focused on communication, justification, and reasoning; proper use of manipulatives; multiple models and representations; and problem-solving strategies. The guide encouraged use of multiple grouping arrangements (individual work, paired instruction, small and large groups).

Tutoring

Tutoring occurred at varying times during the regular school day in the schools students attended, outside the classroom, in the quietest location available, often the library. Students were not pulled out during reading or math instruction. So tutoring was layered on top of the classroom mathematics instructional program. Tutors were full- or part-time employees of the research grant that funded this study.

NC tutoring—We refer to the NC tutoring protocol as Math Flash because NCs “flash” during the computerized practice activity. The Math Flash protocol relies on scripts to (a) clarify for tutors how to frame precise, effective explanations and (b) provide tutors a concrete model for how to implement lessons. Tutors study scripts; they do not read them. Each lesson lasts 20–30 min, and the Math Flash standard protocol runs 16 weeks, with three sessions per week.

Math Flash addresses the 200 NCs with addends and subtrahends from 0 to 9. NCs are introduced in a deliberate order. For the first two lessons, tutors address NCs of +1 and – 1, using manipulatives and the number line, teaching the commutative property of addition, and emphasizing that this property does not apply to subtraction. In the next two lessons, NCs of +0 and –0 are introduced, again with manipulatives and the number line. In Lessons 5 and 6, +1, – 1, +0, and –0 are reviewed.

In Lesson 7, students begin learning doubles from 0 through 6 ($0 + 0$, $1 + 1$, $2 + 2$, $3 + 3$, $4 + 4$, $5 + 5$, $6 + 6$, $0 - 0$, $2 - 1$, $4 - 2$, $6 - 3$, $8 - 4$, $10 - 5$, $12 - 6$) using manipulatives and rehearsing doubles chants. At this point, mastery criteria are introduced, with students spending a minimum of one session on each lesson topic (so students do not waste time on NCs they already know) and a maximum of four sessions on each lesson topic (to prevent students from getting stuck and losing content coverage). Mastery is assessed in each lesson during computerized practice (see below). After doubles, students learn +2 and –2, again using manipulatives and the number line.

Next, students are taught to use two strategies for answering an NC. Students are taught that if they “just know” the NC, then “pull it out of your head.” If, however, they do not know an answer immediately, they count up. Counting-up strategies for addition and subtraction are taught with the number line and the student’s fingers. To count up addition NCs, students start with the bigger number and count up the smaller number on their fingers. The answer is the last number spoken. For subtraction counting up, new vocabulary is introduced. The *minus number* is the number directly after the minus sign. The *number you start with* is the first number in the equation. To count up subtraction NCs, students start with the minus number and count up to the number they start with. The answer is the number of fingers used to count up. From this point on, during every subsequent lesson, students are reminded to “know it or count up.”

Because students are now equipped with two strategies for answering NCs, the tutor introduces additional NC sets beginning with the 5 set. This includes all addition problems equaling 5 and all subtraction problems with 5 as the minuend. After mastery of the 5 set, students progress to the 6 set, then the 7 set, and so on: 8, 9, 10, 11, 12, 13, 14, 15, 16, and 17–18. Students work on each set for a maximum of four sessions. Between the 12 set and the 13 set, students work on doubles of 7–10: $7 + 7$, $8 + 8$, $9 + 9$, $10 + 10$, $14 - 7$, $16 - 8$, $18 - 9$, $20 - 10$. If a student masters all sets before Session 48, the remaining sessions are dedicated to review.

Each of the 48 Math Flash daily lessons comprises five activities: flash card warm-up, conceptual and strategic instruction, lesson-specific flash card practice, computerized practice with mastery assessment, and paper-pencil review. In addition, throughout every lesson, a systematic reinforcement program is used to motivate good attention, hard work, and accurate work.

With flash card warm-up, tutors show flash cards, one at a time for 2 min. These flash cards are a representative sample of the pool of the 200 NCs addressed in Math Flash. Cards answered correctly are placed in a correct pile. When students answer incorrectly, the tutor instructs them to “count up.” Students count up to produce the correct answer, but the card is placed in the incorrect pile. At 2 min, the number of cards answered correctly is counted, and the student graphs this number on a graph.

During conceptual and strategic instruction, tutors introduce or review concepts and strategies. Throughout, tutors emphasize the two strategies for deriving answers (know it and count up), while providing practice in counting up and requiring students to explain how to count up addition and subtraction problems. Tutors then work with students on that session’s NC set (e.g., $+1$ and -1 , doubles 0–6, combinations of 12, etc.) using the number line and manipulatives.

Next, tutors conduct lesson-specific flash card practice for 1 min. Lesson-specific flash cards are the NCs for that session’s lesson (e.g., if a lesson focuses on the 5 set, lesson-specific flash cards are NCs with sums or minuends of 5). Correctly answered cards are placed in the correct pile. When students answer incorrectly, tutors require them to “count up,” and the card is returned to the stack. After 1 min, the number of flash cards answered correctly is counted, but the score is not graphed. On the second, third, and fourth sessions of a lesson topic, students get a chance to beat that session’s lesson-specific flash card score. Tutors remind students what their score on the first minute was and encourage them to do better in the upcoming minute. Scoring and feedback are the same as in the first minute. Tutors praise students when they beat their score.

For the next 7.5 min, students complete computerized practice to build fluency with NCs and to assess mastery with the session’s NC set. NCs presented on the computer include 10 lesson-specific NCs and 5 review NCs. An NC flashes on the screen for 1.3 s. Students rehearse the

NC (e.g., $3 + 2 = 5$) while it briefly appears; when the NC disappears, students retype the entire NC (e.g., addends and answer). If the answer is correct, the student hears applause and earns a point. If the answer is incorrect, the student has another chance to enter the NC correctly. Computerized practice ends after the student answers each of the 10 lesson-specific NCs correctly two times or after 7.5 min. The student then receives performance feedback. Mastery on the lesson-specific NCs set is assessed automatically as the student works on the computer. If the student answers each of the 10 lesson-specific NCs correctly two times before 7.5 min elapse, *mastered* appears on the screen. If not, *review* appears. The tutor moves students on to the next NCs set when mastery occurs or after four sessions on a given set.

Finally, students complete a paper-and-pencil review. The student has 1 min to complete 15 lesson-specific NCs on one side of a paper and then has another minute to complete 15 review NCs on the other side. At 2 min, tutors circle correct answers and write the score at the top of the paper. Students take home these papers each session.

A systematic reinforcement program is incorporated in NC tutoring. Tutors award gold stars following each component of the tutoring session, with the option to withhold stars for inattention or poor effort. Throughout the session, each gold star earned is placed on a “star chart.” Sixteen stars lead to a picture of a treasure box, and when this is reached, the student chooses a small prize from a real treasure box. The student keeps the old star chart and receives a new chart in the next lesson.

WP tutoring—We refer to the WP tutoring protocol as Pirate Math because posters and materials incorporate a pirate theme. Scripts are studied, not read. Each lesson lasts 20–30 min, and the Pirate Math standard protocol runs 16 weeks, with three sessions per week. These 48 lessons are divided into four units. An introductory unit addresses skills foundational to WPs. In this first unit, tutors teach students the counting-up strategy for solving addition and subtraction NCs; review double-digit addition and subtraction; teach students to solve for X in any position in simple algebraic equations (i.e., $a + b = c$; $x - y = z$); and teach students to check their WP work.

The remaining three units focus on WPs, while incorporating and reviewing the foundational skills taught in the introductory unit. Each unit introduces one WP type, and after the first problem-type unit, subsequent units provide systematic, mixed cumulative review that includes previously taught problem types. The WP types are Total (two or more amounts being combined), Difference (two amounts being compared), and Change (initial amount that increases or decreases). See Appendix. In the Total unit, the first problem type taught, tutors teach students to run through a problem: a three-step strategy prompting students to read the problem, underline the question, and name the problem type. Students use the run strategy across all three problem types.

Next, for each problem type, students are taught to identify and circle relevant information. For example, for Total problems, students circle the item being combined and the numerical values representing that item, and then label the circled numerical values as $P1$ (i.e., for Part 1), $P2$ (i.e., for Part 2), and T (i.e., for the combined total). Students mark the missing information with an X and construct an algebraic equation representing the underlying mathematical structure of the problem type. For Total problems, the algebraic equation takes the form of $P1 + P2 = T$, and the X can appear in any of the three variable positions. Students are taught to solve for X , to provide a word label for the answer, and to check the reasonableness and accuracy of work. The strategy for Difference problems and Change problems follows similar steps but uses variables and equations specific to those problem types. For Difference problems, students are taught to look for the bigger amount (labeled B), the smaller amount (labeled s), and the difference between amounts (labeled D) and to use the algebraic equation

$B - s = D$. For Change problems, students are taught to locate the starting amount (labeled St), the changed amount (labeled C), and the ending amount (labeled E); the algebraic equation for Change problems is $St \pm C = E$ (\pm depends on whether the change is an increase or decrease in amount).

Across problem types, explicit instruction to identify transfer features occurs in four ways. First, students are taught that because not all numerical values in WPs are relevant for finding solutions, they should identify and cross out irrelevant information. Second, students are taught to recognize and solve WPs with the missing information in the first or second position. Third, students learn to apply the problem-solving strategies to WPs that involve addition and subtraction with double-digit numbers with and without regrouping. Finally, students are taught to find relevant information for solving WPs in pictographs, bar charts, and pictures. Across the three problem-type units, previously taught problem types are included for review and practice.

After the introductory unit (six lessons), each Pirate Math daily lesson comprises four activities: flash card warm-up, conceptual and strategic instruction, problem-type flash card practice, and paper-pencil review. Also, in every lesson a systematic reinforcement program is used to motivate good attention, hard work, and accurate work.

The first activity, flash card warm-up, is identical to the flash card warm-up used for Math Flash. The second activity, conceptual and strategic instruction, lasts 15–20 min. Tutors provide scaffolded instruction in solving the three types of WPs, along with instruction on identifying and integrating transfer features, using role playing, manipulatives, instructional posters, modeling, and guided practice. In each lesson, students solve three WPs, with decreasing amounts of support from the tutor.

The third activity involves sorting WPs. Tutors read aloud flash cards, each displaying a WP. The student identifies the WP type, placing the card on a mat with four boxes labeled “Total,” “Difference,” “Change,” or “?” Students do not solve WPs; they sort them by problem type. To discourage students from associating a cover story with a problem type, the cards have similar cover stories with varied numbers, actions, and placement of missing information. After 2 min, the tutor notes the number of correctly sorted cards and provides corrective feedback for up to three errors.

In paper-and-pencil review, students have 2 min to complete 10 addition and subtraction NCs and 4 addition and subtraction double-digit computation items, two of which require regrouping. Then, students have 2 min to complete one WP on the back of the paper. Tutors provide corrective feedback and note the number of correct problems on the top of the sheet. Students take home the paper-and-pencil review sheets. The reinforcement program is analogous to Math Flash, except that the student marks the number of gold coins earned in a session on a treasure map. When the student has 16 coins, the student selects a prize from a treasure box.

Tutoring Fidelity and Time

Every tutoring session was audiotaped. Four research assistants independently listened to tapes while completing a checklist to identify the percentage of essential points in that lesson. We sampled 16.8% of tapes such that treatments, research assistants, and lesson types at each site were sampled comparably. In Nashville, the site where the protocols had been developed, the mean percentage of points addressed was 98.1 ($SD = 2.06$) for NC tutoring and 98.4 ($SD = 2.79$) for WP tutoring. In Houston, the mean percentage of points addressed was 99.5 ($SD = 0.47$) for NC tutoring and 99.2 ($SD = 0.68$) for WP tutoring.

Tutors also recorded the duration of each session. In Nashville, tutoring minutes averaged 1,032 ($SD = 85.08$) for NC tutoring and 997 ($SD = 130.25$) for WP tutoring. In Houston, total tutoring minutes averaged 1,155 ($SD = 130.09$) for NC tutoring and 1,158 ($SD = 184.69$) for WP tutoring. Analysis of variance revealed a significant effect for site, $F(1, 82) = 15.68, p < .001$, with more time in Houston than Nashville. The effect for treatment condition was not significant, $F(1, 82) = 0.18, p = .669$; nor was the interaction between treatment and site, $F(1, 82) = 0.27, p = .603$.

Measures

Screening—The calculations screening measure was the Arithmetic subtest of the Wide Range Achievement Test–3 (WRAT; Wilkinson, 1993), in which students have 10 min to complete calculation problems of increasing difficulty. Median reliability is .94 for ages 5–12 years.

The WP screening measure was a 5-item version of a test originally developed by Riley, Greeno, and Heller (1983). The latest version (Jordan & Hanich, 2000) is a 14-item Single-Digit Story Problems measure involving sums or minuends of 9 or less, reflecting change, combine, compare, and equalize relationships. The tester reads each item aloud; students have 30 s to respond and can ask for rereading before the next item. The score is the number of correct answers. Coefficient alpha in a similar sample was .83, and criterion validity with TerraNova (CTB/McGraw-Hill, 1997) Total Math score was .66 in a representative third-grade sample ($N = 777$; Fuchs, Fuchs, Compton, et al., 2006). For screening, we chose the 5 items with the highest item–total correlations and with the highest difficulty ratings in the sample just described. Alpha for the 5 items was .77–.80 (with correct classification vs. the entire test of 86%–88%).

The reading screening measure was the Reading subtest of the WRAT (Wilkinson, 1993), in which students read aloud letters and words until a ceiling is reached. Reliability is .94.

The IQ screening measure was the two-subtest Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). Vocabulary assesses expressive vocabulary, verbal knowledge, memory, learning ability, and crystallized and general intelligence with 37 items; subjects identify pictures and define words. Matrix Reasoning measures nonverbal fluid reasoning and general intelligence with 32 items; subjects select one of five options that best completes a visual pattern. Reliability exceeds .92.

Assessing tutoring effects—We administered these measures immediately before tutoring began (1 month after screening) and then immediately after tutoring ended. To decrease Type I error and enhance construct coverage, we combined measures that tapped the same constructs, grouping four NC measures with a factor score ($M = 0, SD = 1$). The same was done for two procedural calculation measures. (These factor scores were extracted from loadings from a principal factor analysis involving the specified variables. They were not weighted composites.) With the NC factor score, we assessed the effects of NC tutoring on NC outcomes and compared the effects of NC tutoring delivered in NC tutoring against the more simple and time-efficient counting strategy taught in WP tutoring as a foundational skill. The procedural calculation factor score was used in different ways for the two tutoring conditions: For NC tutoring, the outcome assessed transfer of NC skill to situations requiring more complex calculations; for WP tutoring, the outcome assessed the effects of procedural calculation instruction taught as a foundational skill and practiced in the context of solving WPs.

We also had five WP measures. For NC tutoring, we used these measures to assess additional transfer of NC tutoring. For WP tutoring, we used these measures to investigate the development of WP skill as a direct effect of WP tutoring. We did not combine these five

measures because they tapped different aspects of WP tutoring. Two measures assessed the algebraic foundational skills taught: Find X required solving algebraic equations outside WPs; Number Sentences required generating algebraic equations for WPs without calculating solutions. The third measure, Vanderbilt Story Problems, assessed a mix of simple and complex versions of the problem types directly taught as well as transfer of those taught problem types to novel contexts. The final two WP measures, KeyMath and Iowa, assessed a variety of taught and untaught problem types, some with straightforward contexts and others with novel contexts. KeyMath required students to construct responses, whereas Iowa used a multiple-choice response format.

To assess NC learning, we used four subtests of the Grade 3 Math Battery.² Each subtest comprises 25 NCs presented vertically. Students have 1 min to write answers. The score is the number of correct answers. Agreement was assessed on 100% of protocols by two independent scorers; alpha was computed on this sample. Addition Fact Fluency 0–12 comprises addition NCs with sums of 0–12; Subtraction Fact Fluency 0–12, subtraction NCs with minuends of 0–12; Addition Fact Fluency 0–18, addition NCs with sums of 0–18; and Subtraction Fact Fluency 0–18, subtraction NCs with minuends of 0–18. For the four subtests, respectively, percentage of agreement was 98.9, 98.3, 99.9, and 99.3; alpha was .88, .91, .86, and .89.

To assess procedural calculation learning, we used two measures. With Double-Digit Mixed Addition and Subtraction in the Grade 3 Math Battery,³ students have 5 min to complete 20 two-digit addition and subtraction problems with and without regrouping. The score is the number of correct answers. Agreement, calculated on 100% of protocols by two independent scorers, was 99.3%. Alpha on this sample was .93. With Curriculum-Based Measurement–Computation,⁴ students have 3 min to complete 25 addition and subtraction items sampling the typical second-grade curriculum. The score is the number of problems correct. Two independent scorers reentered item-by-item responses into a computerized scoring program on an item-by-item basis. Agreement was 99.7%. Alpha on this sample was .95.

To assess WP learning, we used five measures, none of which included any problem that had been used for instruction. The first two measures, which were administered in small groups and required constructed responses, assess the foundational algebra skills taught in WP tutoring. With Find X,⁵ students solve algebraic equations ($a + b = c$ or $d - e = f$) that vary the position of X across all three slots. The tester demonstrates how to find X with a sample problem. All protocols were independently rescored; agreement was 99.3%. Alpha was .93. With Number Sentences,⁶ the tester reads eight WPs aloud; students have 30 s to write the algebraic equation representing the problem model (students do not find solutions). The score is the number of correct equations. All protocols were independently rescored; agreement was 99.5%. Alpha was .84.

The final three measures assessed students' ability to solve WPs. With Vanderbilt Story Problems,⁷ students complete 18 novel problems representing the three taught problem types (Total, Difference, and Change relationships with missing information in all three positions), with and without irrelevant information and with and without charts or graphs. In small groups, the tester reads each WP aloud; students have 1 min to write a constructed response. Credit is earned for correct math and labels in answers. Alpha was .86. KeyMath–Revised Problem Solving (Connolly, 1998) includes 18 WPs of increasing difficulty, which involve all four

²Available from Lynn S. Fuchs, Vanderbilt University, 228 Peabody, Nashville, TN 37203.

³Available from Lynn S. Fuchs, Vanderbilt University, 228 Peabody, Nashville, TN 37203.

⁴Available from Lynn S. Fuchs, Vanderbilt University, 228 Peabody, Nashville, TN 37203.

⁵Available from Lynn S. Fuchs, Vanderbilt University, 228 Peabody, Nashville, TN 37203.

⁶Available from Lynn S. Fuchs, Vanderbilt University, 228 Peabody, Nashville, TN 37203.

⁷Available from Lynn S. Fuchs, Vanderbilt University, 228 Peabody, Nashville, TN 37203.

operations representing taught and untaught problem types. Administration is one to one; items are read aloud; responses are constructed. Testing is discontinued after three consecutive errors. Split-half reliability at third grade is .72. Correlations with the Total Mathematics score of the Iowa (Hoover, Hieronymous, Dunbar, & Frisbie, 1993) at Grades 1–8 is .60. With the Iowa Test of Basic Skills, for Problem Solving and Data Interpretation (Hoover et al., 1993), students solve 22 WPs representing taught and untaught problem types; data in tables and graphs are required to solve items. The test is administered in small groups, with a multiple-choice response format. At Grades 1–5, Kuder–Richardson 20 is .83–.87.

Procedure

Screening to identify students occurred in September through the use of WRAT–Arithmetic (in large groups) and WPs (in large groups), WRAT–Reading (individually), and WASI (individually). Pretest data were collected in October in small groups (except for one-to-one administration for make-ups and for KeyMath). For all WP measures, the tester read each problem aloud, with opportunities for rereading, and provided enough time for work completion before moving to the next item. Tutoring began the first week of November and ran through the second week of March. Posttesting, which used the pretest measures as well as the Iowa, occurred during the third and fourth week of March. Trained research assistants collected data using standardized directions (except the Iowa was read aloud).

Results

Preliminary Analyses

We conducted distributional exploration of each measure via statistical (e.g., skewness, kurtosis) and graphical (e.g., box plots, stem and leaf plots) means. Generally, the standardized variables were normally distributed at both time points. However, Vanderbilt Story Problems was positively skewed and kurtotic at both time points, and Find X was negatively skewed at pretest and bimodal at posttest. In the case of Vanderbilt Story Problems, a square-root transformation was completed, which improved distribution; however, because results for the original and the transformed variables were similar, results of the original variable are presented for ease of interpretation. Similarly, for Find X, logistic analyses (dichotomizing scores into high and low) yielded similar results as the original variables; so we retained the original form. Number Sentences at pretest and procedural computations at posttest also showed some skewness, but transformations did not generally improve distributions.

Age was unrelated to pre- or posttest performance. Also, tutoring time was unrelated to all but two outcomes, and in this case, the relation was small and did not interact with other effects or change conclusions. Therefore, these variables are not reported. The factors of interest were difficulty status (MD vs. MDRD), site (Houston vs. Nashville), and tutoring condition (NC tutoring vs. WP tutoring vs. control).

Pretest Performance

See Table 2 for pretest tutoring and difficulty status effects. We present pretest site effects in this section. Students in the three treatment groups did not differ on any measure ($p > .05$). For Find X, the only effect was for site, $F(1, 128) = 7.77, p < .006$; Houston outperformed Nashville ($M = 0.24, SD = 1.03$ vs. $M = -0.25, SD = 0.91; d = 0.50$). For KeyMath Problem Solving, Vanderbilt Story Problems, and Number Sentences, there were effects for difficulty status and site; for site: KeyMath, $F(1, 127) = 6.86, p < .01$; Vanderbilt Story Problems, $F(1, 128) = 23.78, p < .0001$; Number Sentences, $F(1, 128) = 16.00, p < .0001$. MD outperformed MDRD. Houston outperformed Nashville (KeyMath, $M = 0.26, SD = 1.00$ vs. $M = -0.25, SD = 0.91, d = 0.53$; Vanderbilt Story Problems, $M = 0.40, SD = 1.04$ vs. $M = -0.41, SD = 0.76, d = 0.88$; Number Sentences, $M = 0.35, SD = 0.95$ vs. $M = -0.36, SD = 0.93, d = 0.79$).

For the NC and Procedural Calculations factors, pretest results were more complicated. For NCs, there were interactions of treatment with site, $F(2, 123) = 5.93, p < .004, \eta_p^2 = .088$, and treatment with difficulty status, $F(2, 123) = 3.43, p < .04, \eta_p^2 = .04$. For NC tutoring and WP tutoring, Houston outperformed Nashville, $F(1, 41) = 13.99, p < .0007$ and $F(1, 39) = 13.70, p < .0007$, respectively, but students assigned to these treatments of different difficulty status did not differ ($p > .05$ for both). Control students across sites did not differ ($p > .05$), but MD outperformed MDRD, $F(1, 44) = 6.01, p < .02$. For Procedural Calculations, there were also interactions of treatment with site, $F(2, 123) = 3.78, p < .03, \eta_p^2 = .090$, and treatment with difficulty status, $F(2, 123) = 3.64, p < .03, \eta_p^2 = .058$. For NC tutoring and WP tutoring, Houston outperformed Nashville, $F(1, 41) = 24.53, p < .0001$, and $F(1, 39) = 12.28, p < .002$, respectively, but students assigned to these treatments of different difficulty status did not differ ($p > .05$ for both). Control students across sites or across difficulty status did not differ ($p > .05$ for both).

Posttest Performance

Given this pattern of pretest results, the design of the study, and the research questions, the primary outcome analysis was three-way analysis of covariance. The factors were treatment (with three levels), difficulty status (with two levels), and site (with two levels). Pretest performance was the covariate. Interactions of these factors with one another and with pretest were examined systematically in reverse order. Highest level interactions were tested first and, when not significant, trimmed from future models. Trimming occurred until a final model was derived, which either retained the highest level of significant interactions (and all lower level interactions) or did not contain interactions; if the latter case pertained, we also evaluated models that retained only the Treatment \times Difficulty Status interaction, and results were not different. All omnibus models were significant, although we do not provide these results to highlight the significant unique contributions. For significant main effects or interactions, we conducted follow-up comparisons on the adjusted posttest means using a correction for multiple comparisons. We also computed effect sizes (Cohen's d), using unadjusted group means in the numerator and the pooled standard deviation across the groups being compared in the denominator to correct for sample overestimation bias (Hedges & Olkin, 1985). Table 3 displays F values for significant effects. Tables 4 and 5, respectively, show posttest means and standard deviations and effect sizes by treatment condition and by difficulty status.⁸ Significant site effects are presented in the text below.

For NCs, there were no interactions among the three factors or with pretest, including the interaction between difficulty status and treatment. In the final model, there were significant effects for pretest, difficulty status, and treatment, but not for site, after controlling for all other factors included in the analysis. MD outperformed MDRD ($p < .035, d = 0.43$). Students in NC tutoring ($p < .0002, d = 0.55$) and WP tutoring ($p < .003, d = 0.62$) outperformed control students, whereas the two tutoring groups did not differ.

For Procedural Calculations, there were no interactions among the three factors or with pretest, including the interaction of difficulty status and treatment. In the final model, there were significant effects for pretest, site, and treatment, after controlling for all other factors. Difficulty status was not significant, but MD students had higher scores ($d = 0.31$). Houston

⁸The impact of socioeconomic status was evaluated via status of subsidized lunch. The proportion of students within the treatments, within sites, within site and treatment, and within difficulty status and treatment did not differ, although students who did not receive subsidized lunch were more likely to be MD than MDRD, as expected. Those who received subsidized lunch did not differ from those who did not, on any measure at pre- or posttest. Adding subsidized lunch to the final models did not change results, and evaluating the moderating impact of subsidized lunch on all the other factors also did not substantively alter interpretation of the treatment effects. In the final models, the inclusion of the classroom nesting component did not alter interpretation of results for any outcome measure.

outperformed Nashville ($M = 0.20$, $SD = 0.98$ vs. $M = -0.20$, $SD = 0.99$; $p < .05$, $d = 0.40$). Students in NC tutoring ($p < .007$, $d = 0.27$) and WP tutoring ($p < .0001$, $d = 0.53$) outperformed control students, but students in the two tutoring groups did not differ.

For Find X, there were no interactions among the three factors or with pretest, including the interaction between difficulty status and treatment group. In the final model, there were significant effects for pretest, site, difficulty status, and treatment, after controlling for all other factors. Houston outperformed Nashville ($M = 0.36$, $SD = 0.81$ vs. $M = -0.37$, $SD = 1.24$; $p < .001$, $d = 0.78$), and MD students outperformed MDRD students ($p < .0003$, $d = 0.69$). WP tutoring students outperformed control students ($p < .04$, $d = 0.36$), whereas NC tutoring students did not differ from either of the other groups.

For Number Sentences, the only interaction was between pretest and difficulty status, in which pretest scores were more strongly related to outcomes of students with MDRD relative to those with MD. There were also significant main effects for pretest, site, and treatment, after controlling for all other factors. There was no main effect for difficulty status, although this was expected given the interaction with pretest. Houston students outperformed Nashville students ($M = 0.40$, $SD = 0.73$ vs. $M = -0.41$, $SD = 1.08$; $p < .0002$, $d = 0.89$). WP tutoring students outperformed control students ($p < .0001$, $d = 0.74$) and NC tutoring students ($p < .0001$, $d = 0.73$); the latter two groups did not differ.

For Vanderbilt Story Problems, there were no interactions among the three factors or with pretest, including the interaction between difficulty status and treatment. In the final model, there were significant effects for pretest, site, difficulty status, and treatment, after controlling for all other factors. Houston outperformed Nashville ($M = 0.34$, $SD = 0.90$ vs. $M = -0.34$, $SD = 0.99$; $p < .013$, $d = 0.71$), and MD students outperformed MDRD students ($p < .013$, $d = 0.62$). WP tutoring students outperformed the control group ($p < .0001$, $d = 0.79$) and NC tutoring students ($p < .0001$, $d = 0.83$); the latter two groups did not differ.

For KeyMath Problem Solving, there were no interactions among the three factors or with pretest, including the interaction between difficulty status and treatment. In the final model, there were significant effects for pretest, site, difficulty status, and treatment, after controlling for all other factors. Houston outperformed Nashville ($M = 0.53$, $SD = 0.94$ vs. $M = -0.56$, $SD = 1.00$; $p < .0001$, $d = 1.31$), and MD students outperformed MDRD students ($p < .0001$, $d = 1.29$). WP tutoring students outperformed control students ($p < .03$, $d = 0.28$), whereas NC tutoring students did not differ from either of the other groups.

The Iowa was the only measure administered exclusively at posttest. Therefore, the other commercial WP measure administered at pretest, KeyMath, was used as the pretest covariate. There were no interactions among the three factors or with pretest, including the interaction between difficulty status and treatment. In the final model, there were significant effects for pretest, site, and difficulty status, but not for treatment, after controlling for all other factors. Houston students outperformed Nashville students ($M = 0.38$, $SD = 1.03$ vs. $M = -0.41$, $SD = 0.77$; $p < .0003$, $d = 0.86$), and MD students outperformed MDRD students ($p < .031$, $d = 0.85$).

Discussion

We assessed the efficacy of two tutoring protocols for remediating deficits among third-grade students with mathematics difficulty. Results demonstrated the efficacy of these tutoring protocols for remediating key deficits of third-grade students with mathematics difficulty, with robust effects across two urban sites that varied in important ways. Despite variations in proximity to the treatment developers, in mathematics programs, and in students' level of mathematics competence, results revealed no significant interaction between site and treatment on any implementation or outcome measure. Tutors at the two sites implemented the protocols

comparably well, and students at the two sites responded to those protocols comparably well. These findings provide the basis for concluding that the tutoring protocols are transportable. It suggests the potential for scaling up these protocols, given the proviso that tutors are trained as in the present study: with one session of instruction; with practice implementing the procedures alone and with each other during the subsequent week; with a practice session conducted with a supervisor who provides corrective feedback; with tutors studying (not reading) scripts; and with meetings among tutors and the supervisor every 2–3 weeks to address problems or questions as they arise.

With respect to fluency with NCs, both tutoring conditions effected superior improvement compared with the control group, with no significant difference between the tutoring conditions. Compared with the control group, the effect size for NC tutoring was 0.55, and the effect size for WP tutoring was similar (0.62). The comparability of outcomes for the two tutoring conditions is notable because NC tutoring allocated dramatically more time to NCs over the 16-week intervention. With NC tutoring, each 20–30-min session was devoted entirely to NCs. By contrast, with WP tutoring, tutors taught a counting strategy for deriving NC solutions in a single lesson, and then provided practice each session with the 2-min warm-up activity, with 2 min of paper–pencil review, and as NC errors occurred in WPs. On this basis, we conclude that teaching students an efficient counting strategy, while providing frequent but small amounts of timed practice to gain efficiency in using that strategy and while contextualizing the use of that strategy within WPs, effects comparable outcomes to an expanded tutoring protocol devoted entirely to NCs. This raises questions about the added value of the conceptual lessons about NCs or the more extensive drill and practice in the NC tutoring condition. Given the efficiency of the WP-embedded counting strategies remediation, it may be the remediation of choice for this population of students with mathematics difficulties who experience counting deficits, immature counting strategies, and slow retrieval times (Fleischner et al., 1982; Geary et al., 1987, 1992, 2007; Goldman et al., 1988). Future work should systematically vary approaches to NC remediation to extend present findings.

In terms of procedural calculations, both tutoring conditions again effected superior outcomes compared with the control group. In this case, the effect size compared with the control condition was 0.27 for NC tutoring but almost double that for WP tutoring (0.53). This difference was not statistically significant; yet, on the basis of these effect sizes, we speculate that with larger samples, WP tutoring might achieve differential efficacy compared with NC tutoring. Such a finding would not be surprising because only WP tutoring allocated direct, albeit limited, time to procedural calculations (with one direct lesson conducted in the introductory foundational skills unit, with 2 min of paper–pencil practice at the end of each session, and with students completing procedural calculations while solving WPs).

It is, however, notable that even without direct work on procedural calculations, NC tutoring effected better outcomes on procedural calculations compared with the control group, indicating that transfer occurred. This is theoretically important because NCs are viewed as a signature, bottleneck deficit for students with mathematics disability (Fleischner et al., 1982; Geary et al., 1987; Goldman et al., 1988). As explained earlier in this article, the hypothesis is that with a fixed amount of attention, students with NC deficits allocate available resources to deriving answers to these simple problems instead of focus on the demands of the more complex mathematics into which NCs are embedded (cf. Ackerman et al., 1986; Goldman & Pellegrino, 1987). If NCs represent a signature deficit, performance on more complex mathematics tasks should improve simply as a function of NC remediation, just as decoding intervention has been shown to improve reading comprehension (Blachman et al., 2004; Torgesen et al., 2001). We found support for this hypothesis in the transfer we observed from NC remediation to procedural calculation outcomes, suggesting that NCs may in fact serve as a bottleneck deficit, at least with respect to procedural calculations.

However, we found no evidence to support this hypothesis on WP outcomes. With NC improvement (but in the absence of WP tutoring), students with mathematics difficulties evidenced no WP improvement. This suggests that the source of their difficulty is not diverting attention from the complex mathematics to the NCs embedded in those problems, but rather failing to comprehend the relations among the numbers embedded in the narratives or to process the language in those stories adequately. This suggests that NCs are not the bottleneck for WP performance. Instead, it indicates that mathematics disability represents a more complicated pattern of difficulty, implicating language as has been suggested elsewhere (e.g., Fuchs et al., 2005; Fuchs, Fuchs, Compton, et al., 2006). Given these contradictory findings about transfer from NC remediation, in which NC remediation transferred to procedural calculations but not to WPs, future work should continue to explore this issue focusing on WPs as well as other components of the mathematics curriculum.

It is important to emphasize that WP tutoring was efficacious. First, it promoted algebraic skill related to WPs. On Find X, in which students solve three-number algebraic equations that vary the operation and the position of X, only WP tutoring effected superior outcome compared with the control group. The effect size was 0.36. On Number Sentences, in which students generate algebraic equations to represent WP models without solving the equations, again only WP tutoring effected superior outcome compared with the control group. This time, the effect size was 0.74. In the case of Number Sentences, WP tutoring students also outperformed NC tutoring students. Although this is not surprising given that only WP tutoring addressed these skills and that algebra is otherwise completely novel at third grade, these results show that the algebraic cognition of the WP tutoring students improved as a function of tutoring that incorporates algebra as a tool for solving WPs. After all, these students were not only severely deficient in incoming math skill but also young. Given the strong focus on algebra in high schools and the requirement in many states that students pass an algebra course or test prior to graduation, introducing algebra this early in the curriculum may represent a productive innovation.

Importantly, work on these foundational skills (NCs, procedural calculations, and algebra), combined with the schema-broadening instruction provided in WP tutoring, not only effected improvement on NCs, procedural calculations, and algebra but also produced differential growth on WP outcomes. As might be expected, effects favoring WP tutoring were more dramatic and consistent when WP measures were better aligned with tutoring (even though none of the problems on any outcome measures had been used for instruction). Vanderbilt Story Problems included only problems representing the three problem types directly addressed in WP tutoring (Total, Difference, and Change relationships), with missing information in all three positions, with and without irrelevant information, and with and without charts or graphs. The response format was also consistent with instruction, requiring constructed responses. On this measure, WP tutoring effected substantially superior outcomes. This was true not only when compared against the control group (with an effect size of 0.79) but also when compared against NC tutoring (with an effect size of 0.83). KeyMath Problem Solving was a more distal measure because, although students responded with constructed responses as in tutoring, the measure assessed taught as well as untaught problem types. On KeyMath, WP tutored students outperformed the control group (with a smaller effect size of 0.28); however, the two tutoring conditions achieved comparably. The most distal measure, the Iowa, not only included a variety of taught and untaught problem types but also required multiple-choice responses. On this measure, there were no significant effects.

In this way, across a range of WP measures, findings support the efficacy of the Pirate Math WP tutoring protocol to enhance the WP outcomes among students with serious math difficulties. At the same time, findings suggest the importance of aligning WP instruction with high-stakes outcome measures by addressing the complete set of problem types assessed on

those measures and by incorporating the constructed response format (as would occur in school and in real life) as well as the multiple-choice response formats that appear on high-stakes tests.

Besides assessing the efficacy of the tutoring protocols, another major purpose of the present study was to examine whether tutoring is differentially efficacious depending on students' difficulty status: MD versus MDRD. Because a key deficit among students with reading difficulty is phonological processing and because phonological processing deficits are linked with difficulty in automatic retrieval of NCs (see Geary, 1993), we hypothesized that MDRD students would be less responsive to NC tutoring than MD students. In addition, because using text to construct a WP model involves language (e.g., Fuchs et al., 2005; Fuchs, Fuchs, Compton et al., 2006; Fuchs, Fuchs, Stuebing et al., 2008; Swanson, 2006; Swanson & Beebe-Frankenberger, 2004) and because the language profiles of students with MDRD are depressed compared with those of students with MD (Powell et al., in press), we hypothesized that MDRD students would be less responsive to WP tutoring than MD students.

Yet, we found no evidence of differential responsiveness to intervention as a function of difficulty status on any outcome: None of the interactions between treatment condition and difficulty status were significant. This raises questions about the tenability of the MD–MDRD subtyping scheme and suggests the need to pursue other avenues for subtyping mathematics disability (Fletcher et al., 2007). For example, some work (Fuchs, Fuchs, Stuebing et al., 2008) suggests that calculations disability versus WP disability may represent a productive subtyping framework. Even so, across tutoring conditions (and sites), students with MD did outperform students with MDRD at pre- and posttest. Additional work to examine the tenability of the MD–MDRD subtyping scheme is warranted, even as research pursuing alternative frameworks proceeds.

In the meantime, the complete absence of interactions between treatment and difficulty status in the present study indicates that the main effects favoring tutoring conditions apply across students with mathematics difficulty, regardless of their reading skill. In sum, NC tutoring (i.e., Math Flash) enhances automatic retrieval of NCs with transfer to procedural calculations but without transfer to algebra or WPs. By contrast, for a comparable amount of tutoring time, WP tutoring (i.e., Pirate Math) that also incorporates instruction on foundational skills (NCs, procedural calculations, and algebra) enhances WP skill as well as fluency with NCs, procedural calculations, and algebra. These findings appear robust, applying across MD and MDRD students and across sites.

Acknowledgments

The research described in this article was supported in part by National Institute of Child Health and Human Development (NICHD) Core Grant P30HD15052 to Vanderbilt University and by NICHD Award P01046261 to the University of Houston and through subcontract to Vanderbilt University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NICHD or the National Institutes of Health.

References

- Ackerman PT, Anhalt JM, Dykman RA. Arithmetic automatization failure in children with attention and reading disorders: Associations and sequelae. *Journal of Learning Disabilities* 1986;19:222–232. [PubMed: 3701218]
- Andersson U, Lyxell B. Working memory deficit in children with mathematical difficulties: A general or specific deficit? *Journal of Experimental Child Psychology* 2007;96:197–228. [PubMed: 17118398]
- Ashcraft MH, Stazyk EH. Mental addition: A test of three verification models. *Memory & Cognition* 1981;9:185–196.
- Badian, NA. Dyscalculia and nonverbal disorders of learning. In: Myklebust, HR., editor. *Progress in learning disabilities*. Vol. 5. New York: Grune & Stratton; 1983. p. 235-264.

- Blachman BA, Schatschneider C, Fletcher JM, Francis DJ, Clonan SM, Shaywitz BA, Shaywitz SE. Effects of intensive reading remediation for second and third graders and a 1-year follow-up. *Journal of Educational Psychology* 2004;96:444–461.
- Boaler J. Encouraging the transfer of “school” mathematics to the “real world” through the integration of process and content, context, and culture. *Educational Studies in Mathematics* 1993;25:341–373.
- Bransford, JD.; Schwartz, DL. Rethinking transfer: A simple proposal with multiple implications. In: Iran-Nejad, A.; Pearson, PD., editors. *Review of research in education*. Vol. 24. Washington, DC: American Educational Research Association; 1999. p. 61-100.
- Brown, AL.; Campione, JC.; Webber, LS.; McGilly, K. Interactive learning environments: A new look at assessment and instruction. In: Gilford, BR.; O’Connor, MC., editors. *Changing assessments: Alternative views of aptitude, achievement, and instruction*. Boston: Kluwer Academic; 1992. p. 37-75.
- Bruck M. Persistence of dyslexics’ phonological awareness deficits. *Developmental Psychology* 1992;28:874–886.
- Chi MTH, Feltovich PJ, Glaser R. Categorization and representation of physics problems by experts and novices. *Cognitive Science* 1981;5:121–152.
- Christensen CA, Gerber MM. Effectiveness of computerized drill and practice games in teaching basic math facts. *Exceptionality* 1990;1:149–165.
- Cirino PT, Ewing-Cobbs L, Barnes M, Fuchs LS, Fletcher JM. Cognitive arithmetic differences in learning disability groups and the role of behavioral inattention. *Learning Disabilities Research and Practice* 2007;22:25–35.
- Clements DH, Sarama J. Effects of a preschool mathematics curriculum: Summative research on the Building Blocks Project. *Journal for Research in Mathematics Education* 2007;38:136.
- Compton DL, Fuchs LS, Fuchs D. The course of reading and mathematics disability in first grade: Identifying latent class trajectories and early predictors. 2007Manuscript submitted for publication
- Connolly, AJ. *KeyMath-Revised, Normative Update: A diagnostic inventory of essential mathematics*. Circle Pines, MN: American Guidance Service; 1998.
- Cooper G, Sweller J. Effects of schema acquisition and rule automation on mathematical problem solving transfer. *Journal of Educational Psychology* 1987;79:347–362.
- CTB/McGraw-Hill. *TerraNova technical manual*. Monterey, CA: Author; 1997.
- Fleischner JE, Garnett K, Shepherd MJ. Proficiency in arithmetic basic fact computation of learning disabled and nondisabled children. *Focus on Learning Problems in Mathematics* 1982;4:47–56.
- Fletcher, JM.; Lyon, GR.; Fuchs, LS.; Barnes, MA. *Learning disabilities: From identification to intervention*. New York: Guilford; 2007.
- Foxman, D.; Ruddock, G.; McCallum, I.; Schagen, I. *APU mathematics monitoring (Phase 2)*. Slough, England: National Foundation for Educational Research; 1991.
- Fuchs LS, Compton DL, Fuchs D, Paulsen K, Bryant JD, Hamlett CL. The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology* 2005;97:493–513.
- Fuchs LS, Fuchs D. Mathematical problem solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities* 2002;35:563–574. [PubMed: 15493253]
- Fuchs LS, Fuchs D, Compton DL, Powell SR, Seethaler PM, Capizzi AM, et al. The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology* 2006;98:29–43.
- Fuchs LS, Fuchs D, Craddock C, Hollenbeck KN, Hamlett CL, Schatschneider C. Effects of small-group tutoring with and without validated classroom instruction on at-risk students’ math problem solving: Are two tiers of prevention better than one? *Journal of Educational Psychology* 2008;100:491–509. [PubMed: 19122881]
- Fuchs LS, Fuchs D, Finelli R, Courey SJ, Hamlett CL. Expanding schema-based transfer instruction to help third graders solve real-life mathematical problems. *American Educational Research Journal* 2004;41:419–445.
- Fuchs LS, Fuchs D, Finelli R, Courey SJ, Hamlett CL, Sones EM, Hope SK. Teaching third graders about real-life mathematical problem solving: A randomized controlled study. *Elementary School Journal* 2006;106:293–312.

- Fuchs LS, Fuchs D, Prentice K. Responsiveness to mathematical problem-solving instruction among students with risk for mathematics disability with and without risk for reading disability. *Journal of Learning Disabilities* 2004;4:293–306. [PubMed: 15493402]
- Fuchs LS, Fuchs D, Prentice K, Burch M, Hamlett CL, Owen R, et al. Explicitly teaching for transfer: Effects on third-grade students' mathematical problem solving. *Journal of Educational Psychology* 2003;95:293–304.
- Fuchs LS, Fuchs D, Prentice K, Hamlett CL, Finelli R, Courey SJ. Enhancing mathematical problem solving among third-grade students with schema-based instruction. *Journal of Educational Psychology* 2004;96:635–647.
- Fuchs LS, Fuchs D, Stuebing K, Fletcher JM, Hamlett CL, Lambert WE. Problem solving and calculation skill: Shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology* 2008;100:30–47.
- Fuchs LS, Fuchs D, Yazdian L, Powell SR. Enhancing first-grade children's mathematical development with peer-assisted learning strategies. *School Psychology Review* 2002;31:569–584.
- 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.
- Gagne RM. Contributions of learning to human development. *Psychological Review* 1968;75:177–191. [PubMed: 4874111]
- Geary DC. Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin* 1993;114:345–362. [PubMed: 8416036]
- Geary DC, Bow-Thomas CC, Yao Y. Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology* 1992;54:372–391. [PubMed: 1453139]
- Geary DC, Brown SC, Samaranayake VA. Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology* 1991;27:787–797.
- Geary DC, Hamson CO, Hoard MK. Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology* 2000;77:236–263. [PubMed: 11023658]
- Geary DC, Hoard MK, Byrd-Craven J, Nugent L, Numtee C. Cognitive mechanisms underlying achievement deficits in children with mathematics learning disability. *Child Development* 2007;78:1343–1359. [PubMed: 17650142]
- Geary DC, Widaman KF, Little TD, Cormier P. Cognitive addition: Comparison of learning disabled and academically normal elementary school children. *Cognitive Development* 1987;2:249–269.
- Glaser RM. Education and thinking: The role of knowledge. *American Psychologist* 1983;39:94–104.
- Goldman SR, Pellegrino JW. Information processing and educational microcomputer technology: Where do we go from here? *Journal of Learning Disabilities* 1987;20:249–269.
- Goldman SR, Pellegrino JW, Mertz DL. Extended practice of addition facts: Strategy changes in learning-disabled students. *Cognition and Instruction* 1988;5:223–265.
- Greenes, C.; Larson, M.; Leiva, MA.; Shaw, JM.; Stiff, L.; Vogeli, BR.; Yeatts, K. Houghton Mifflin math. Boston: Houghton Mifflin; 2005.
- Griffin, SA.; Case, R.; Siegler, RS. Rightstart: Providing the central conceptual prerequisite for first formal learning of arithmetic to students at risk for school failure. In: McGilly, K., editor. *Classroom lessons: Integrating cognitive theory and classroom practice*. Cambridge, MA: MIT Press; 1994. p. 25–50.
- Groen GJ, Parkman JM. A chronometric analysis of simple addition. *Psychological Review* 1972;79:329–343.
- Gross-Tsur V, Manor O, Shalev RS. Developmental dyscalculia: Prevalence and demographic features. *Developmental Medicine and Child Neurology* 1996;37:906–914.
- Hanich LB, Jordan NC, Kaplan D, & Dick J. Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology* 2001;93:615–626.
- Hedges, LV.; Olkin, I. *Statistical methods for meta-analysis*. Orlando, FL: Academic Press; 1985.

- Hoover, HD.; Hieronymous, AN.; Dunbar, SB.; Frisbie, DA. Iowa Test of Basic Skills, Form K. Itasca, IL: Riverside; 1993.
- Jitendra AK, DiPipi CM, Perron-Jones N. An exploratory study of schema-based word-problem-solving instruction for middle school students with learning disabilities: An emphasis on conceptual and procedural understanding. *Journal of Special Education* 2002;36:23–38.
- Jitendra AK, Griffin CC, Haria P, Leh J, Adams A, Kaduvettoor A. A comparison of single and multiple strategy instruction on third-grade students' mathematical problem solving. *Journal of Educational Psychology* 2007;99:115–127.
- Jitendra AK, Griffin CC, McGoey K, Gardill MC, Bhat P, Riley T. Effects of mathematical word problem solving by students at risk or with mild disabilities. *Journal of Educational Research* 1998;91:345–355.
- Jitendra AK, Hoff K. The effects of schema-based instruction on the word-problem-solving performance of students with learning disabilities. *Journal of Learning Disabilities* 1996;29:421–431.
- Jordan NC, Hanich L. Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities* 2000;33:567–578. [PubMed: 15495398]
- Jordan NC, Hanich LB, Kaplan D. Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology* 2003;85:103–119. [PubMed: 12799164]
- Jordan NC, Montani TO. Cognitive arithmetic and problem solving: A comparison of children with specific and general mathematics difficulties. *Journal of Learning Disabilities* 1997;30:624–634. [PubMed: 9364900]
- Kilpatrick, J.; Swafford, J.; Findell, B., editors. *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press; 2001.
- Larkin, JH. What kind of knowledge transfers?. In: Resnick, LB., editor. *Knowing, learning, and instruction: Implications for reform*. Hillsdale, NJ: Erlbaum; 1989. p. 197-208.
- Mawer R, Sweller J. What do students learn while solving mathematics problems? *Journal of Educational Psychology* 1985;77:272–284.
- Micallef S, Prior M. Arithmetic learning difficulties in children. *Educational Psychology* 2004;24:175–200.
- Miller SP, Hudson PJ. Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge. *Learning Disabilities Research and Practice* 2007;22:47–57.
- Okolo CM. The effect of computer-assisted instruction format and initial attitude on the arithmetic facts proficiency and continuing motivation of students with learning disabilities. *Exceptionality* 1992;3:195–211.
- Ostad SA. Developmental differences in addition strategies: A comparison of mathematically disabled and mathematically normal children. *British Journal of Educational Psychology* 1997;67:345–357. [PubMed: 9376311]
- Powell SR, Fuchs LS, Fuchs D, Cirino P, Fletcher JM. Word-problem performance as a function of math disability, with and without reading disability. *Journal of Learning Disabilities*. in press
- Quilici JL, Mayer RE. Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology* 1996;88:144–161.
- Reikeras EKL. Performance in solving arithmetic problems: A comparison of children with different levels of achievement in mathematics and reading. *European Journal of Special Needs Education* 2006;21:233–250.
- Resnick, LB.; Resnick, DP. Assessment the thinking curriculum: New tools for educational reform. In: Gilford, BR.; O'Connor, MC., editors. *Changing assessments: Alternative views of aptitude, achievement, and instruction*. Boston: Kluwer Academic; 1992. p. 37-75.
- Riley, MS.; Greeno, JG.; Heller, JI. Development of children's problem-solving ability in arithmetic. In: Ginsburg, HP., editor. *The development of mathematical thinking*. Orlando, FL: Academic Press; 1983. p. 153-196.
- Rivera-Batiz FL. Quantitative literacy and the likelihood of employment among young adults in the United States. *Journal of Human Resources* 1992;27:313–328.
- Siegler RS. The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General* 1987;116:250–264.

- Swanson HL. Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of Educational Psychology* 2006;98:265–281.
- Swanson HL, Beebe-Frankenberger M. The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology* 2004;96:471–491.
- Torgesen JK, Alexander AW, Wagner RK, Rashotte CA, Voeller KS, Conway T. Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities* 2001;34:33–58. [PubMed: 15497271]
- Tournaki N. The differential effects of teaching addition through strategy instruction versus drill and practice to students with and without learning disabilities. *Journal of Learning Disabilities* 2003;36:449–458. [PubMed: 15497488]
- Wechsler, D. Wechsler Abbreviated Scale of Intelligence. San Antonio, TX: Psychological Corporation; 1999.
- Wilkinson, GS. Wide Range Achievement Test 3. Wilmington, DE: Wide Range; 1993.
- Xin PX, Jitendra AK, Deatline-Buchman A. Effects of mathematical word problem-solving instruction on middle school students with learning problems. *Journal of Special Education* 2005;39:181–192.

Table 1

Demographic and Screening Data by Tutoring Condition and Difficulty Status

Variable	Tutoring condition			Difficulty status		
	WPs (<i>n</i> = 42)	NCs (<i>n</i> = 44)	Control (<i>n</i> = 47)	MD (<i>n</i> = 73)	MDRD (<i>n</i> = 60)	
Age (in years)	8.98 (0.6)	9.00 (0.6)	8.86 (0.5)	8.73 (0.5) ^a	9.19 (0.5) ^b	
Female	45%	52%	34%	40%	48%	
Subsidized lunch	76%	82%	77%	68%	90%	
Special education	17%	18%	17%	8%	28%	
Retained	29%	30%	23%	15%	19%	
English as second language	19%	14%	15%	12%	20%	
African American	57%	61%	70%	66%	60%	
Caucasian	7%	11%	9%	11%	7%	
Hispanic	26%	25%	19%	19%	28%	
Other	10%	3%	2%	4%	5%	
WASI full-scale IQ	87.71 (11.9)	89.41 (9.8)	88.86 (12.3)	91.86 (11.8) ^a	84.80 (9.4) ^b	
WRAT-Arithmetic	83.52 (10.1)	84.66 (11.0)	85.55 (8.8)	87.71 (9.0) ^a	80.85 (9.9) ^b	
WRAT-Reading	91.43 (18.7)	91.59 (13.6)	94.68 (15.8)	104.64 (7.2) ^a	78.02 (11.0) ^a	

Note. Percentages are computed relative to the number of individuals in that group (e.g., 45% of the 42 students in Pirate Math tutoring were female). Percentages for ethnicity within a column total 100%. Numbers in parentheses are standard deviations. For mean scores, values in a given row with different subscripts are significantly different from one another. WPs = word problems; NCs = number combinations; MD = math difficulty; MDRD = math and reading difficulty; WASI = Wechsler Abbreviated Scale of Intelligence; WRAT = Wide Range Achievement Test-3.

Table 2

Pretest Performance by Tutoring Condition and Difficulty Status

Variable	Tutoring condition				Difficulty status		
	<i>F</i> (<i>df</i>)	WPs (<i>n</i> = 42)	NCs (<i>n</i> = 44)	Control (<i>n</i> = 47)	<i>F</i> (<i>df</i>)	MD (<i>n</i> = 73)	MDRD (<i>n</i> = 60)
Number combinations	(2, 123), <1	0.10 (0.92)	-0.12 (1.10)	0.02 (0.98)	(1, 123), <1	0.11 (0.99)	-0.13 (1.01)
Procedural calculations	(2, 123), <1	-0.07 (0.94)	-0.06 (1.05)	0.11 (1.02)	(1, 123), <1	0.12 (1.00)	-0.14 (0.99)
Word problems							
Find X	(2, 128), <1	-0.09 (1.02)	-0.02 (0.96)	0.10 (1.03)	(1, 128), <1	0.06 (1.04)	-0.07 (0.95)
Number Sentences	(2, 128), 1.10	-0.15 (0.99)	-0.04 (1.02)	0.17 (0.98)	(1, 128), 6.41*	0.23 (0.98)	-0.29 (0.95)
KeyMath	(2, 127), <1	-0.06 (1.13)	-0.02 (0.89)	0.07 (0.99)	(1, 127), 4.91*	0.20 (1.02)	-0.25 (0.92)
Vanderbilt Story Problems	(2, 128), 2.06	0.13 (1.11)	-0.21 (0.84)	0.08 (1.02)	(1, 128), 7.07*	0.25 (1.06)	-0.30 (0.84)

Note. *F* values are for the main effects indicated, after controlling for other factors in the model; for number combinations (NCs) and procedural calculations, there were interactions of tutoring group and/or difficulty status with each other and/or with site. Scores in table are raw means (uncorrected). NCs and procedural calculations are factor scores derived from a combination of similar measures. KeyMath, Vanderbilt Story Problems, and Find X, however, are individual pretest variables, which have been *z* score standardized for comparative purposes. The Iowa was not administered at pretest because it would have produced a floor effect at that time. WPs = word problems; MD = math difficulty; MDRD = math and reading difficulty.

* $p < .05$.

Table 3

F Values at Posttest

Outcome measures	Effects			
	Pretest (1)	Difficulty status (1)	Treatment (2)	Site (1)
Number combinations (127)	89.99***	4.56*	10.63***	NS
Procedural calculations (127)	161.02***	NS	13.10***	4.06*
Word problems				
Find X (127)	16.56***	13.91***	3.22*	11.32***
Number sentences (117) ^a	15.50***	NS	17.48***	14.30***
KeyMath (126)	36.16***	16.76***	3.36*	45.83***
Iowa (126)	43.47***	4.77*	NS	13.93***
Vanderbilt Story Problems (127)	19.45***	6.47*	14.64***	6.39*

Note. Denominator degrees of freedom are in parentheses next to the measures; numerator degrees of freedom are next to the effects. Denominator degrees of freedom change because different effects were trimmed from the model depending on which interactions occurred for which outcomes. NS = not significant.

^aThe interaction between pretest and difficulty was significant, $F(1, 117) = 7.06, p < .01$.

* $p < .05$.

*** $p < .001$.

Table 4

Posttest Performance by Tutoring Condition and Difficulty Status

Variable	Tutoring condition			Difficulty status	
	WPs (<i>n</i> = 42)	NCs (<i>n</i> = 44)	Control (<i>n</i> = 47)	MD (<i>n</i> = 73)	MDRD (<i>n</i> = 60)
Number combinations	0.20 (0.94)	0.19 (1.10)	-0.36 (0.86)	0.19 (0.96)	-0.23 (1.00)
Procedural calculations	0.29 (0.92)	0.01 (0.89)	-0.26 (1.11)	0.14 (0.97)	-0.17 (1.02)
Word problems					
Find X	0.21 (0.84)	-0.03 (1.01)	-0.15 (1.10)	0.30 (0.77)	-0.36 (1.13)
Number Sentences	0.51 (1.15)	-0.23 (0.83)	-0.24 (0.84)	0.20 (0.94)	-0.25 (1.02)
KeyMath	0.14 (1.05)	-0.02 (0.99)	-0.14 (0.94)	0.55 (0.98)	-0.46 (0.81)
Iowa	0.12 (1.15)	-0.12 (0.84)	-0.03 (0.97)	0.25 (0.98)	-0.35 (0.90)
Vanderbilt Story Problems	0.57 (1.27)	-0.28 (0.67)	-0.25 (0.76)	0.27 (1.06)	-0.33 (0.82)

Note. *F* values are for the main effects indicated, after controlling for other factors in the model; for number combinations (NCs) and procedural calculations, there were interactions of tutoring group and/or difficulty status with each other and/or with site. Scores in table are raw means (uncorrected). NCs and procedural calculations are factor scores derived from a combination of similar measures. KeyMath, Vanderbilt Story Problems, and Find X, however, are individual pretest variables, which have been *z* score standardized for comparative purposes. The Iowa was not administered at pretest because it would have produced a floor effect at that time. WPs = word problems; MD = math difficulty; MDRD = math and reading difficulty.

Effect Sizes by Tutoring Condition and Difficulty Status

Variable	Contrast tutoring condition				Contrast difficulty status: MD vs. MDRD
	WPs vs.				
	Control	NCs	NCs vs. Control		
Number combinations	+0.62 (0.19 to 1.04)	+0.01 (-0.41 to 0.43)	+0.55 (0.14 to 0.97)	+0.43 (0.08 to 0.77)	
Procedural calculations	+0.53 (0.11 to 0.96)	+0.31 (-0.12 to 0.73)	+0.27 (-0.15 to 0.68)	+0.31 (-0.03 to 0.65)	
Word problems					
Find X	+0.36 (-0.06 to 0.78)	+0.26 (-0.17 to 0.68)	+0.11 (-0.30 to 0.52)	+0.69 (0.34 to 1.04)	
Number Sentences	+0.74 (0.31 to 1.18)	+0.73 (0.30 to 1.17)	+0.01 (-0.40 to 0.41)	+0.46 (0.11 to 1.04)	
KeyMath	+0.28 (-0.14 to 0.70)	+0.16 (-0.27 to 0.58)	+0.12 (-0.29 to 0.53)	+1.29 (0.91 to 1.66)	
Iowa	+0.14 (-0.28 to 0.56)	+0.24 (-0.19 to 0.66)	-0.10 (-0.51 to 0.31)	+0.85 (0.50 to 1.21)	
Vanderbilt Story Problems	+0.79 (0.36 to 1.22)	+0.83 (0.39 to 1.28)	-0.04 (-0.45 to 0.37)	+0.62 (0.27 to 0.97)	

Note. Effect sizes are based on unadjusted means at posttest and pooled standard deviation of both groups being compared. Effect sizes for difficulty status appear larger than suggested in Table 3 because significance (though not actual means) in Table 3 is adjusted for treatment group. Positive effect sizes indicate higher means for the first group named in the comparison. Values in parentheses are confidence intervals for the effect sizes. WPs = word problems; NCs = number combinations; MD = math difficulty; MDRD = math and reading difficulty.

Appendix

Sample Problems by Problem Type and by Position of Missing Information ($A + B = C$ or $A - B = C$)

Problem	Problem type	Position of missing information
The art teacher had 82 pieces of colored paper. Some of the pieces of paper were blue, and 20 of the pieces were green. How many pieces of blue paper did she have?	Total	A ($x + 20 = 82$)
Donna and Natasha made 96 friendship bracelets. Donna made 25 bracelets. How many friendship bracelets did Natasha make?	Total	B ($25 + x = 96$)
There are 51 boys and 47 girls in the third grade at Baker Elementary School. How many third graders are there?	Total	C ($51 + 47 = x$)
Charles put 14 more roses than daisies in the vase. He put 25 daisies in the vase. How many roses did he put in the vase?	Difference	A ($x - 25 = 14$)
Maurice has 11 more comic books than Thomas. Maurice has 37 comic books. How many comic books does Thomas have?	Difference	B ($37 - x = 11$)
At the picnic, the kids ate 65 hot dogs. They ate 32 hamburgers. How many more hot dogs did they eat than hamburgers?	Difference	C ($65 - 32 = x$)
The temperature outside this morning was cool. By the afternoon, the temperature had gone up 35 degrees so it is now 87 degrees outside. What was the temperature in the morning?	Change	A ($x + 35 = 87$)
Jamarius baked 78 chocolate chip cookies. Then he gave some to his friends. Now Jamarius has 23 cookies. How many cookies did Jamarius give to his friends?	Change	B ($78 - x = 23$)
Mr. Luther had 26 pencils in his desk. Then he gave 12 pencils to his students. How many pencils does Mr. Luther have now?	Change	C ($26 - 12 = x$)

Note. Other problems included irrelevant information or incorporated relevant information in pictographs, scenes, or bar charts. Students only see the problem. They are taught to solve the problem by identifying the problem type; generating an algebraic sentence to represent that problem type; and solving for x within the algebraic equation.