*Learn Individ Differ*. Author manuscript; available in PMC 2011 April 1.

Published in final edited form as:

*Learn Individ Differ*. 2010 April 1; 20(2): 89–100. doi:10.1016/j.lindif.2009.09.003.

# The Effects of Strategic Counting Instruction, with and without Deliberate Practice, on Number Combination Skill among Students with Mathematics Difficulties

**Lynn S. Fuchs**,
Vanderbilt University

**Sarah R. Powell**,
Vanderbilt University

**Pamela M. Seethaler**,
Vanderbilt University

**Paul T. Cirino**,
University of Houston

**Jack M. Fletcher**,
University of Houston

**Douglas Fuchs**, and
Vanderbilt University

**Carol L. Hamlett**
Vanderbilt University

## Abstract

The primary purpose of this study was to assess the effects of strategic counting instruction, with and without deliberate practice with those counting strategies, on number combination (NC) skill among students with mathematics difficulties (MD). Students ($n$ = 150) were stratified on MD status (i.e., MD alone vs. MD with reading difficulty) and site (proximal vs. distal to the intervention developer) and then randomly assigned to control (no tutoring) or 1 of 2 variants of NC remediation. Both remediations were embedded in the same validated word-problem tutoring protocol (i.e., Pirate Math). In 1 variant, the focus on NCs was limited to a single lesson that taught strategic counting. In the other variant, 4–6 min of practice per session was added to the other variant. Tutoring occurred for 16 weeks, 3 sessions per week for 20–30 min per session. Strategic counting without deliberate practice produced superior NC fluency compared to control; however, strategic counting with deliberate practice effected superior NC fluency and transfer to procedural calculations compared with both competing conditions. Also, the efficacy of Pirate Math word-problem tutoring was replicated.

Inquiries should be sent to Lynn S. Fuchs, 228 Peabody, Vanderbilt University, Nashville, TN 37203 (615-343-4782; lynn.fuchs@vanderbilt.edu).

**Keywords**

strategic counting; number combinations; math facts; intervention; mathematics difficulties; randomized experiment

## 1. Introduction

Mathematics disability creates life-long challenges (Rivera-Batiz, 1992) for 5–9% of the population (e.g., Badian, 1983; Shalev, Auerbach,, Manor, & Gross-Tsur, 1996). Although early prevention activities can substantially reduce prevalence (e.g., Fuchs, Compton, Fuchs, Paulsen, Bryant, & Hamlett, 2005), no prevention program is universally effective. Therefore, the need for remedial intervention persists even when strong prevention services are available. In the present study, we focused on the remediation of number combination (NC) deficits. We begin by discussing the importance of NC skill and how it develops in students with and without mathematics disability. Then, we summarize previous work on the remediation of NC deficits. Finally, we explain how the present study extends the literature.

### 1.1 The Importance and Development of NC Skill

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), with evidence that fluency with NCs is a significant path to procedural computation and word-problem performance (Fuchs, Fuchs, Compton et al., 2006). In developing competence with NCs, a common pattern of development involves children gradually gaining efficiency with counting strategies. When adding, children initially count two sets (e.g., 2+3) in their entirety (i.e., 1, 2, 3, 4, 5). This is referred to as the *sum counting strategy* (or counting all). As they discover the number-after principle reflected in counting, they realize that the sum of 5+1 is the number that occurs after 5 when counting (Baroody, 1995). Fluent number–after knowledge makes 1+n/n+1 NCs among the easiest to learn (Baroody). The rule for adding 1 provides developmental scaffold for more abstract, sophisticated counting-on strategies: Once children recognize that the sum of 5+1 is the number after 5 in the counting sequence, they understand that the sum of 5+2 cannot be 6 but instead is two numbers past 5: 6, 7 (Baroody). In this way, children discover the efficiency of counting up from the first addend. When the smaller addend occurs first, as in 2+3, they begin counting from the smaller addend (i.e., 2, 3, 4, 5), using the *max strategy*. Eventually, as they understand the commutative property of addition, they discover the most efficient *min counting strategy* – or counting from the larger addend (i.e., 3, 4, 5), regardless of whether the larger addend appears first or second.

As conceptual knowledge about number grows, children discover the additive identity property of zero and can therefore answer n+0/0+n NCs. Also, children learn that a whole can be decomposed into parts in different ways, and this big idea sets the stage for decomposition strategies to derive answers (e.g., 2+3=[2+2=4]+1=5). As increasingly efficient counting and decomposition strategies help students consistently and quickly pair problems with correct answers in working memory, associations become established in long-term memory, and children gradually favor memory-based retrieval of answers (Ashcraft & Stazyk, 1981; Geary, Widaman, Little, & Cormier, 1987; Goldman et al., 1988; Groen & Parkman, 1972; Siegler, 1987).

Eventually, in developing a formal part-whole understanding of addition, students discover the meaning of missing-addend expressions (e.g., for 2+?=5: 2 is the known part; ? is the unknown part; 5 indicates the whole, which must be larger than either part). This promotes skill with addition NCs and helps children discover the missing-addend counting strategy for subtraction (e.g., 5−3: 3 is the known part; 4, 5 is the unknown part) (Baroody, 1995). Also, as children develop understanding of the relationship between subtraction and addition (Baroody, 1999; Baroody, Ginsburg, & Waxman, 1983), knowledge of addition NCs facilitates knowledge of subtraction NCs.

Research (e.g., Ashcraft & Stazyk, 1981; Geary et al., 1987; Goldman et al., 1988; Groen & Parkman, 1972; Siegler, 1987) documents that competent NC performance involves a mix of strategies, with counting strategies and decomposition strategies serving as back-ups for primary reliance on memory-based retrieval, and that even adults, use varying strategies at different times to solve the same NC. Although the extent to which counting strategies, decomposition strategies, and memory-based retrieval are explicitly addressed in general education varies (Miller & Hudson, 2007), typical students nonetheless are relatively adept with NCs by the end of third grade (Cirino, Ewing-Cobbs, Barnes, Fuchs, & Fletcher, 2007).

Students with mathematics disability manifest greater difficulty with counting (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Geary, Bow-Thomas, & Yao, 1992); persist with immature back-up strategies (Geary et al., 2007); and fail to make the shift to memory-based retrieval of answers (Fleishner, Garnett, & Shepherd, l982; Geary et al., l987; Goldman et al., l988). 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, l991; Shalve et al., 2000; Ostad, 1997). Some (e.g., Fleishner et al.; Geary et al., 1987; Goldman et al.) consider NCs to be a signature deficit of students with mathematics disability, and deficits in 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).

### 1.2 Previous Work on Remediating NC Deficits

Three major approaches for remediating NC deficits involve providing drill and practice, developing conceptual understanding to foster decomposition strategies, and teaching strategic counting. Okolo (1992) and Christensen and Gerber (1990) contrasted computerized drill and practice in a game versus a drill format. Okolo found no significant differences, whereas Christensen and Gerber found that students were disadvantaged by the game. Tournaki (2003) contrasted paper-pencil drill and practice with a strategic counting condition. Results showed an advantage for strategic counting, but the strategic counting condition incorporated strong instructional principles whereas the drill and practice condition did not. In addition, these prior studies failed to include a control, making it unclear whether remediation leads to better progress than would be expected with business-as-usual schooling. Also, the authors did not specify whether mathematics was a domain of impairment among these students with school-identified learning disabilities, making it unclear whether effects apply to students who experience mathematics difficulty (MD).

We extended this literature in a series of studies by including a control group while using random assignment and screening participants to ensure mathematics difficulty. In the first study (Fuchs, Powell, Hamlett, Fuchs, Cirino, & Fletcher, 2008), our approach to remediation was drill and practice, although instead of simply requiring students to answer NCs, we engineered software to ensure that students practiced correct responses. Results showed that on NC outcomes, students who received this NC drill and practice outperformed those in the three competing conditions, with large effect sizes (0.69–0.78). We concluded that reliably and efficiently pairing question stems with correct responses was efficacious, but questioned whether a stronger focus on developing conceptual understanding to foster decomposition strategies might enhance learning.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Consequently, in Powell et al. (2009), we randomly assigned students to four conditions, two of which focused on NCs: drill and practice as in Study 1 and drill and practice as in Study 1 plus explicit conceptual instruction focused largely on decomposition strategies. Across the four conditions, students did not differ at pretest. Despite more instructional time for the latter intervention, the two NC conditions resulted in comparable effects (which were reliably stronger than the two competing conditions). This suggested that explicit conceptual instruction to help students develop decomposition strategies for solving NCs may not impart added value over our approach to intensive drill and practice.

In our next study (Fuchs, Powell, Seethaler, Fuchs, Cirino, & Fletcher, 2009), we turned our attention to teaching strategic counting, which involves teaching students the efficient counting procedures (i.e., min for addition; missing addend for subtraction). Although students are not explicitly taught strategic counting in school, typically developing students, unlike students with MD, discover these strategies on their own (Ashcraft & Stazyk, 1981; Geary et al., 1987; Goldman et al., 1988; Groen & Parkman, 1972; Siegler, 1987). Because of inefficient counting strategies, MD students pair question stems with answers slowly, taxing short-term memory, and their answers are often incorrect. Long-term representations for automatic retrieval of NCs therefore fail to establish correctly. It is also possible that students with MD have special difficulty committing NCs to memory. We hoped that explicit instruction on counting strategies would build NC fluency, even if students failed to rely on automatic retrieval for their major strategy. We contrasted two conditions that incorporated strategic counting. One combined strategic counting with intensive drill and practice (as in Studies 1 and 2). The other, which was embedded in word-problem remediation, taught the same strategic counting but practice was confined to the counting strategies for 4–6 min each session. On NC outcomes, both remediations effected superior NC fluency compared to the control group (effects sizes: 0.52 and 0.58). The comparability of outcomes for the two remediation groups is notable, because strategic counting with elaborated drill and practice allocated dramatically more time to NCs over the course of 42 (of 48) intervention sessions sessions in which practice occurred: 20–30 min per session versus 4–6 min per session.

### 1.3 Purpose of the Present Study

The primary purpose of the present study was to assess the effects of strategic counting instruction (in which students are taught the min strategy for addition and the missing addend strategy for subtraction), with and without deliberate practice with those counting strategies, on NC fluency among students with MD. We contrasted a no-tutoring control group against two variants of strategic counting instruction. Both were embedded in word-problem remediation. In one variant, the focus on NCs was limited to a single lesson that simply taught the counting strategies (i.e., strategic counting instruction without deliberate practice). In the other variant, students were taught counting strategies in the same single lesson, but then also practiced strategic counting for answering NCs for 4–6 min each session (i.e., strategic counting instruction with deliberate practice).[1] Pinpointing the value of practice in this controlled way is important because although many assume its necessity for students with MD (e.g., Carnine, 1997;Coyne et al., 2009;Jitendra et al., 2001), no studies have isolated its effects for this population of learners.

We were also interested in investigating whether the effects of strategic counting remediation, with and without deliberate practice, differ as a function of whether MD occurs alone (MD)

[1]The present study differs from Fuchs et al. (2009) in that in the following ways. First, in the present study, both conditions were embedded within word-problem tutoring and the focus on NC remediation was limited to counting strategies; by contrast, in Fuchs et al., one condition was entirely devoted to NC remediation and also incorporated computerized NC practice and other forms of NC drill and practice. Second, in the present study, we had a novel contrast that was not examined by Fuchs et al. That is, we contrast strategic counting instruction (confined to one lesson) with and without deliberate practice.

or in combination with reading difficulty (MDRD). Geary (1993) hypothesized that MD versus MDRD may be a productive scheme for subtyping mathematics disability, because a key deficit associated with reading difficulty is phonological processing (e.g., Bruck, 1992) and because phonological processing deficits are linked to difficulty with automatic retrieval of NCs (e.g., Fuchs et al., 2005). For these reasons, MDRD students should experience greater difficulty with NCs compared to students with MD. Compared to students with MDRD, those with MD use more efficient counting procedures to solve NCs (Geary, Hamson, & Hoard, 2000; Jordan & Hanich, 2000) with faster retrieval times (Andersson & Lyxell, 2007; Hanich, Jordan, Kaplan, & Dick, 2001; Jordan & Montani, 1997) but with comparable accuracy (Cirino et al., 2007). However, the literature is not consistent (e.g., Micallef & Prior, 2004; Reikeras, 2006), and most studies have employed cross-sectional causal-comparative designs. In the present study, we used an experimental approach with random assignment to provide the basis for stronger, causal inferences about the tenability of the subtyping scheme. Based on the causal-comparative literature showing slower retrieval times for MDRD (e.g., Cirino et al.) and implicating phonological processing in the development of NC skill (Fuchs et al.), we hypothesized an interaction between treatment condition and MD status whereby students with MD would respond comparably to the two active treatments, but students with MDRD would respond better with deliberate practice.

As mentioned, in the present study, we embedded the counting strategies with and without deliberate practice conditions within word-problem tutoring. The word-problem tutoring protocol, referred to as Pirate Math because students find X using algebraic equations to represent the structure of word problems, has been shown to be efficacious in enhancing word-problem skill in two prior randomized control trials, with effect sizes ranging from .28 to 1.80 (Fuchs et al., 2009; Fuchs, Seethaler et al., 2008). Pirate Math's word-problem instruction is based on schema theory (e.g., Brown et al., 1992; Cooper & Sweller, 1987). A *schema* is a category that encompasses similar problems; it is a problem type (Chi, Feltovich, & Glaser, l981; Quilici & Mayer, l996). The broader the category for the problem type (i.e., the broader the schema), the greater the probability 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, the tutor first teaches problem-solution rules for a problem type. Then the tutor helps students broaden their schema for this problem type and their awareness of those schemas (Cooper & Sweller). Pirate Math incorporates schema-broadening instruction in these ways.

## 2. Method

### 2.1 Participants

The study was conducted in two large urban school districts, in Nashville (where the tutoring protocols were developed) and Houston (distal site). Third-grade students (*n*=1682) were screened for inclusion in 13 Nashville schools and 18 Houston schools. Because both active tutoring protocols focused on NCs as well as word problems, students were initially screened on a calculations measure and on a word-problem measure. The criterion for low performance was < 26$^{th}$ percentile (i.e., below average) on calculations and < 36$^{th}$ percentile on word problems, criteria that are similar to those used in the MD literature (e.g., Geary et al., 2007; Mazzocco, 2007). Of the 1682 students, 279 (17%) met both criteria. These students were further screened on a reading and an IQ measure. We excluded 56 students who scored between the 25$^{th}$ and 40$^{th}$ percentiles in reading; six students with a T-score below 30 on both IQ subtests; and one student for both reasons. Finally, 36 students met both math criteria (word problems and calculations) for study entry, but were excluded due to limited English proficiency or because they moved away before treatment assignment or because the study had already met its resource limit before they were assigned. This left 180 students eligible to participate.

Students scoring < 26$^{th}$ percentile on the reading measure were classified as MDRD (*n* = 136); those scoring > 39$^{th}$ percentile on the reading were classified as having MD (*n* = 44). The 180 students (90 at each site) represented 45 classrooms in Nashville and 39 classrooms in Houston. 860 students (51%) scored above cutoffs on both measures, whereas 543 students (32%) met criteria for one but not the other (18% met calculations only; 14% met problem solving only). Despite the higher cutoff level for word problems, fewer students actually met this criterion alone, relative to those who met the calculations criterion alone. Blocking on site and difficulty status (MD vs. MDRD), we randomly assigned students to three treatment conditions: strategic counting with deliberate practice (*n* = 60), strategic counting without deliberate practice (*n* = 61), and control (*n* = 59). The two counting strategies conditions were embedded in Pirate Math word-problem tutoring. Because of the design, the composition of treatment conditions was similar on site and difficulty status, $p > .05$.

Thirty students (17%) moved before posttesting: 11 in Nashville and 19 in Houston (nine from the strategic counting with deliberate practice, 12 from the strategic counting without deliberate practice, and nine from the control group). Analyses proceeded with the 150 students with posttest data. The mean age of this sample at mid-year was 8.37 (*SD* = 0.51). Seventy-nine students were in Nashville; 71 in Houston. Forty-one (27%) met criteria for MD; 109 for MDRD. Fifty-one students were assigned to strategic counting with deliberate practice, 49 to strategic counting without deliberate practice, and 50 to the control group. The amount of tutoring in the active treatments was approximately 22 hours (1315 min, *SD* = 169). Twenty-nine students (19%) were classified as English learners; 85 (57%) were boys; 49 (33%) had been retained in at least one grade; 113 (75%) were eligible for subsidized lunch (data were missing for two); 53 (35%) were classified as special education. Eighty-four (56%) were African American; 44 (29%) Hispanic; 16 (11%) Caucasian; and the remaining six students were of other ethnicity (one did not have data). Again, as expected given the design, the treatment conditions in the final sample were comparable on site or difficulty status, $p > .05$. The treatment conditions were also comparable on sex, ethnicity, subsidized lunch status, whether they had been retained in at least one grade, and English as a second language status (all $p > .05$). Both active tutoring groups were however more likely to receive special education relative to the control group, $\chi^2(2, n{=}150) = 9.90$, $p < .008$, creating a conservative test of treatment efficacy. Table 1 provides demographic data by treatment condition and difficulty status.

## 2.2 Classroom Mathematics Program

At both sites, NC instruction in classrooms was minimal. Word problem instruction addressed the three problem types taught in Pirate Math word-problem tutoring as well as more complex problem types. Problem types were taught one at a time and focused on underlying concepts and solution strategies. There was no attempt to broaden students' schemas. In Nashville, the classroom program was *Houghton Mifflin Math* (Greenes et al., 2005). A prescribed set of problem-solution rules was taught, using 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 Pirate Math word-problem tutoring, classroom instruction provided more practice in applying problem-solution rules and provided 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 Texas 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).

### 2.3 Tutoring

Tutoring occurred on a one-to-one basis at varying times during the regular school day in the schools students attended, outside the classroom, in the quietest location available, often the library. Tutoring did not occur during reading or math instruction (except in isolated cases or because of scheduling changes) so tutoring was layered over classroom instruction. Tutors were full- or part-time employees of the research grant that funded this study. Every tutoring lesson was scripted, but scripts were studied, not read. Tutoring ran 16 weeks, with 48 sessions (three per week), with each session lasting 20–30 min. These 48 sessions were divided into four units.

The introductory unit addressed skills foundational to word problems. Tutors taught the single lesson on strategic counting for deriving answers to NCs; reviewed algorithms for answering double-digit addition and subtraction procedural calculations; taught methods to solve for “X” in any position in simple algebraic equations (i.e., a+b=c; x−y=z); and taught strategies for checking work within word problems. The remaining three units focused on word problems, which required the foundational skills taught in the introductory unit.

In the single strategic counting lesson, students were taught that if they “just know” the answer to a NC, they “pull it out of your head.” If, however, they do not know an answer immediately, they “count up.” Strategic counting for addition and subtraction were introduced with the number line. The counting strategy for addition NCs was the min strategy: Students start with the bigger number and count up the smaller number on their fingers. The answer is the last number spoken. The counting strategy for subtraction NCs was the missing addend strategy, which required new vocabulary. The *minus number* is the number directly after the minus sign. The *number you start with* is the first number in the equation. 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.

In the *strategic counting instruction with deliberate practice* condition, practice in strategic counting was incorporated into every lesson in the three word-problem units. The tutor began each lesson by asking the student, “What are the two ways to find an answer to a math fact?” The student responded, “Know it or count up.” Then the tutor asked the student to explain how to count up an addition problem and how to count up a subtraction problem and required the student to count up two addition and two subtraction problems. The next activity in each session was flash card warm up. Students had 1 min to answer NCs. If they responded incorrectly, the tutor required them to count up until they had the correct answer. At the end of 1 min, the tutor counted the cards, and the student then had another min to beat the first score. Also, throughout the lesson, whenever the student made a NC error, the tutor required the student to count up. Finally, when checking the paper-pencil review, the tutor corrected NC errors by demonstrating the counting strategy.

In the *strategic counting instruction without deliberate practice* condition, strategic counting was taught using the same single lesson in the same introductory unit. The lesson was never repeated or reviewed after that single session in the introduction unit, and following this lesson, practice in counting up did not occur. The flash card warm-up activity asked students to read numerals (0–9999) aloud for 1 min. If the student made an error, the tutor asked the student to try again. At the end of 1 min, the tutor counted the correct cards and encouraged the student to beat that score in the upcoming min. Also, throughout the lesson, tutors corrected NC errors by providing the correct answer rather than requiring the student to count up. Finally, when checking the paper-pencil review, the tutor corrected NC errors by simply providing correct answers.

Other than these differences concerning practice with the counting strategies, instruction in the two conditions was identical. The three word-problem units each focused on one word problem type and, after the first problem-type unit, subsequent units provided systematic, mixed cumulative review that included previously taught problem types. The word problem types were Total (two or more amounts being combined), Difference (two amounts being compared), and Change (initial amount that increases or decreases). See Appendix A for sample problems by problem type.

Each word-problem session comprised six activities for the strategic counting instruction with deliberate practice condition and five activities for the strategic counting instruction without deliberate practice condition. The extra activity for strategic counting instruction with deliberate practice was a daily brief review of counting strategies (see above). Otherwise, activities occurred as follows.

With *flash card warm-up*, students had 1 min to answer problems presented on flash cards. At the end of 1 min, the cards were counted, and the student then had another min to beat the first score. This was conducted in parallel fashion in the two active conditions except that the strategic counting with deliberate practice condition focused on NCs, and tutors required students to correct errors by counting up. By contrast, the strategic counting instruction without deliberate practice condition focused on reading numerals, and tutors corrected errors by providing the correct answer.

*Word-problem warm-up* was the next activity; it lasted approximately 2 min and was initiated during. the seventh tutoring session. The tutor showed the student the word problem from the previous day's paper-and-pencil review. The student would explain to the tutor how the word problem was solved.

The next activity, *conceptual and strategic instruction*, lasted 15 to 20 min. Tutors provided scaffolded instruction in the underlying structure of and in solving the three types of word problems (i.e., developing a schema for each problem type), along with instruction on identifying and integrating transfer features (to broaden students' schema for each problem type), using role-playing, manipulatives, instructional posters, modeling, and guided practice. In each lesson, students solved three word problems, with decreasing amounts of support from the tutor. When NC errors occurred, tutors required students in the strategic counting with practice condition to count up; they provided correct responses to students in the strategic counting without practice condition.

In the Total unit, the first problem type covered, tutors taught students to RUN through a problem: a 3-step strategy prompting students to Read the problem, Underline the question, and Name the problem type. Students used the RUN strategy across all three problem types. Next, for each problem type, students were taught to identify and circle the relevant information for that problem type (i.e., schema). For example, for Total problems, students circled the item being combined and the numerical values representing that item, and then labeled the circled numerical values as "P1" (i.e., for part one), "P2" (i.e., for part two), and "T" (i.e., for the combined total). Students marked the missing information with an "X" and constructed an algebraic equation representing the underlying mathematical structure of the problem type. For Total problems, the algebraic equation took the form of "P1 + P2 = T," and the "X" can appear in any of the three variable positions. Students were 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 followed similar steps but used variables and equations specific to those problem types. For Difference problems, students were 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 were 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 +/− C = E” (+/− depends on whether the change is an increase or decrease in amount).

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

*Sorting word problems* was the next activity. Tutors read aloud flash cards, each displaying a word problem. The student identified the word problem type, placing the card on a mat with four boxes labeled “Total,” “Difference,” “Change,” or “?.” Students did not solve word problems; they sorted them by problem type. To discourage students from associating a cover story with a problem type, the cards use similar cover stories with varied numbers, actions, and placement of missing information. After 2 min, the tutor noted the number of correctly sorted cards and provided corrective feedback for up to three errors.

In *paper-and-pencil review*, the final activity, students had 2 min to complete 9 number sentences asking the student to find X. Then, students had 2 min to complete one word problem. Tutors provided corrective feedback and noted the number of correct problems on the paper. In the strategic counting instruction with deliberate practice condition, tutors required students to count up NC errors; in strategic counting instruction without deliberate practice, tutors provided answers. Tutors kept the paper-and-pencil review sheets each day to review the word problem with the student on the following day.

A systematic *reinforcement* program was incorporated in both active conditions. Tutors awarded gold coins following each activity of the session, with the option to withhold coins for inattention or poor effort. Throughout the session, each gold coin earned was placed on a “Treasure Map.” Sixteen coins led to a picture of a treasure box and, when reached, the student chose a small prize from a real treasure box. The student kept the old Treasure Map and received a new map in the next lesson.

### 2.4 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 17.47% of tapes such that treatments, research assistants, and lesson types at each site were sampled comparably. At the site where the protocols had been developed (Nashville), the mean percentage of points addressed was 98.85 (*SD*=1.78) for the strategic counting instruction with deliberate practice condition and 99.09 (*SD*=0.87) for the strategic counting instruction without deliberate practice condition. In Houston, the mean percentage of points addressed was 96.82 (*SD*=3.75) for the strategic counting instruction with deliberate practice condition and 96.43 (*SD*=3.73) for the strategic counting instruction without deliberate practice condition.

Tutors also recorded the duration of each session. Analysis of variance examined tutoring time as a function of site and treatment for the 150 students who completed the year. There was no interaction between site and treatment, although there were significant effects for site, $F(1,97) = 9.11$, $p < .003$, with more time in Houston than Nashville, and for treatment, $F(1,97) = 4.75$, $p = < .04$, with more time for word problem tutoring with practice. In Nashville, tutoring minutes averaged 1,286.61 ($SD = 101.09$) for word problem tutoring with practice, and 1,246.63 ($SD = 131.40$) for word problem tutoring without practice. In Houston, tutoring minutes averaged 1,412.24 ($SD = 197.29$) for word problem tutoring with practice, and 1,312.84 ($SD = 190.23$) for word problem tutoring without practice. However, within the active tutoring groups, time in minutes was unrelated to posttest outcomes (median $r = -0.09$, all $p > .05$), so it was not considered as a covariate in future analyses.

### 2.5 Measures

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

The word-problem screening measure was the Iowa Test of Basic Skills: Problem Solving and Data Interpretation (Iowa; Hoover, Hieronymous, Dunbar, & Frisbie, 1993), where students solve 22 word problems; data in tables and graphs is required to solve items. The test has different test items at different grades. The test is administered in small groups, with a multiple-choice response format. At grades 1–5, KR20 is .83–.87.

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

The IQ screening measure was the 2-subtest Wechsler Abbreviated Scales 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 1 of 5 options that best completes a visual pattern. Reliability exceeds .92.

**2.52 Assessing tutoring effects—**We administered these measures soon after screening and just before tutoring began, and then again immediately after tutoring ended. Because many of the measures were experimental, for ease of comparability, all outcomes were standardized to have a mean of 0 and *SD* of 1. However, raw score means and *SD*s are also available (see Results). Final results did not differ whether raw scores or standardized scores were analyzed.

**NC learning:** We used four subtests of the Grade 3 Math Battery (Fuchs, Powell, & Hamlett, 2003). 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 from 0–12; Subtraction Fact Fluency 0–12, subtraction NCs with minuends from 0–12; Addition Fact Fluency 0–18, addition NCs with sums from 0–18. Subtraction Fact Fluency 0–18, subtraction NCs with minuends from 0–18. For the four subtests, respectively, percentage of agreement was 99.2, 98.7, 99.5, and 98.2; alpha was .87, .92, .87, and .88. To decrease Type I error, we combined the four measures with a factor score (mean = 0; $SD = 1$).

**Procedural calculations learning:** We used Double-Digit Mixed Addition and Subtraction in the Grade 3 Math Battery (Fuchs et al., 2003). Students have 5 min to complete 20 2-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 99.1%. Alpha on this sample was .97.

**Algebraic learning:** We used two measures that assess the foundational algebra skills taught in word-problem tutoring. With Find X (Fuchs & Seethaler, 2008), students solve algebraic equations (a+b=c or d−e=f) that vary the position of X across all 3 slots. The tester demonstrates how to find X with a sample problem. All protocols were independently rescored; agreement was 99.5%. Alpha was .92. With Number Sentences (Fuchs & Seethaler, 2008), the tester reads eight word problems aloud; students have 30 sec to write the algebraic equation representing the problem model (they do not find solutions). The score is the number of correct equations. All protocols were independently re-scored; agreement was 99.7%. Alpha was .82.

**Word-problem learning:** We used three measures: a second- and a third-grade version of Vanderbilt Story Problems (Fuchs & Seethaler, 2008) and KeyMath-Revised Problem Solving (Connolly, 1998). With each version of Vanderbilt Story Problems, students complete 18 novel problems that assess a mix of simple and complex versions of the problem types directly taught (total, difference, and change relationships with missing information in all three positions), as well as transfer of those taught problem types to novel contexts (with and without irrelevant information and with and without charts/graphs). In small groups, the tester reads each word problem aloud; students have 1 min to write a constructed response. Credit is earned for correct math and labels in answers. For the second- and third-grade version, respectively, alpha was .85 and .87. KeyMath includes 18 word problems of increasing difficulty, which involve all four 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. the correlation with the Total Mathematics score of the Iowa (Hoover et al., 1993) is .60.

### 2.6 Procedure

Screening to identify students began in September using WRAT-Arithmetic (in groups) and Iowa (in groups), followed by WRAT-Reading (individually) and WASI (individually). Pretest data were collected soon thereafter individually. For Number Sentences and for all word-problem measures, the tester read each problem aloud, with opportunities for re-reading, and provided enough time for work completion before moving to the next item. Tutoring then occurred three days per week, for 16 weeks. Posttesting occurred at the conclusion of tutoring individually using the pretest measures. Given the timing of the school years, each stage of the study generally occurred about 1 month earlier in Nashville than Houston. Trained research assistants collected data using standardized directions.

## 3.0 Results

### 3.1 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. The screening variables were negatively skewed, which was expected because low performance was required for study entry. Generally, the raw and standardized variables used to assess treatment effects were normally distributed at pre- and posttest. However, for Find X, over one-third of the sample scored zero at pretest, and the distribution was somewhat bimodal at posttest; therefore logistic analyses, dichotomizing scores into low (0–3) and high (4–8) were used to evaluate this outcome. For Number Sentences, pretest distributions were similar to that of Find X (one-third scored zero), but were more evenly distributed at posttest; therefore, this outcome was evaluated with pretest as a dichotomous predictor (low=0–7; high=8–16) but with the posttest as continuous. The raw score for Grade 3 Vanderbilt Story Problems exhibited some positive skew at posttest. A square-root transformation improved its distribution, and this transformation was standardized and evaluated.

### 3.2 Pretest Performance

See Table 2 for pretest performance by treatment condition and difficulty status. For NCs, there was an interaction of site, difficulty status, and treatment, $F(2,136) = 3.71$, $p < .03$, $\eta_p^2$=.052; however, examination of within-site effects revealed no significant interactions of difficulty status and treatment within either site. There was a significant main effect of site, $F(1,143) = 15.58$, $p < .0001$, with Houston students outperforming Nashville students. For Grade 2 Vanderbilt Story Problems, there was an interaction of treatment with difficulty status, $F(2,138) = 3.23$, $p < .05$, $\eta_p^2$=.045. For MD students, the treatment groups were comparable. However, MDRD control students outperformed those in strategic counting with deliberate practice ($p < .05$); MDRD students in strategic counting without deliberate practice did not differ from either other group.

For the remaining pretest variables, there were no interactions among the factors. For the WRAT-3 Arithmetic screening measure and for KeyMath Problem Solving, there were effects for difficulty status, $F(1,145) = 15.08$, $p < .0001$, and $F(1,145) = 11.11$, $p < .001$, respectively, with MD students outperforming those with MDRD; there were no effects for treatment condition or site. For Iowa, there was an effect for site, $F(1,145) = 6.19$, $p < .02$, with Houston students scoring higher than Nashville students, and for treatment, $F(2,145) = 3.28$, $p < .04$, with control students outperforming those assigned to either of the active treatments (both $p < .05$). Using dichotomized pretest performance, a similar pattern was evident for Find X (site: *Wald* $\chi^2$(df=1, N=150) = 27.55, $p < .0001$; treatment: *Wald* $\chi^2$(df=1, N=150) = 6.73, $p < .04$), with Houston students outperforming Nashville students, and students without deliberate practice outperforming those with deliberate practice, $p < .05$. For Vanderbilt Story Problems (Grade 3) and procedural calculations, there were no differences among treatment groups or for difficulty status, although Houston students scored higher than Nashville students, $F(1,145) = 7.71$, $p < .007$, $F(1,145) = 12.59$, $p < .0005$, respectively. There was a similar pattern for Number Sentences, using dichotomized performance, with no significant effect for differences or for difficulty status, but with an effect for site, *Wald* $\chi^2$(df=1, N=150) = 19.33, $p < .0001$ (Houston students scored higher than Nashville students).

Main effects for site or difficulty status do not threaten the validity of study findings, because they apply across the treatment conditions. Main effects favoring Houston over Nashville have been documented previously (e.g., Fuchs et al., 2009) and suggest stronger math instruction at that site. Main effects favoring MD over MDRD are expected based on previous work (e.g., Andersson & Lyxell, 2007; Geary et al., 2000; Hanich et al., 2001; Jordan & Hanich, 2000; Jordan & Montani, 1997). The differences involving treatment, on Grade 2 Vanderbilt Story Problems and Iowa, favor the control group (and echoes the greater proportion of special education students in the active treatments). This creates a conservative test of treatment efficacy. Moreover, pretest performances were accounted for in all subsequent analyses.

### 3.3 Posttest Performance

**3.3.1 Analyses—**The main analysis for evaluating outcomes was a 3-way ANCOVA. The three factors were treatment condition (with 3 levels), difficulty status (with 2 levels), and site (with 2 levels). Pretest performance was the covariate (with additional covariates considered where appropriate). 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, were trimmed from future models. This continued until a model remained with either 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 by difficulty status interaction, and results did not differ. Where there were significant main effects or interactions, follow-up comparisons were conducted on the adjusted posttest means using a correction for multiple comparisons.

In terms of additional covariates, treatment conditions differed on the proportion of students in special education. This was unrelated to posttest variables except for procedural calculations, $F(1,148) = 8.64$, $p < .004$, where we included special education status as a covariate. Age was unrelated to posttest variables, except for KeyMath ($r = -.37$, $p < .0001$), where we included age as a covariate. Subsidized lunch status was unrelated to posttest variables except for KeyMath, $F(1,146) = 8.86$, $p < .003$, and Number Sentences, $F(1,146) = 5.42$, $p < .03$, where we included subsidized lunch status as a covariate. (Tutoring time was considered, but it was unrelated to any outcome.)

Effect sizes (Cohen's *d*) were computed using unadjusted group means in the numerator and the pooled *SD* 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 for the primary factors. Tables 4 and 5 displays posttest means and *SD*s, and Table 6 shows effect sizes by treatment condition and by difficulty status.

**3.3.2 NC learning—**For NCs, there were no significant interactions. Considering other factors, there were main effects for pretest, $p < .0001$, for difficulty status, $p < .0006$, and for treatment condition, $p < .0001$, although not for site. MD students outperformed MDRD students ($d = +0.55$). Students in both active treatments outperformed those in control (for strategic counting with deliberate practice, $d = +0.67$; for strategic counting without deliberate practice $d = +0.43$, both $p < .003$); students with deliberate practice outperformed those without deliberate practice ($d = +0.22$, $p < .03$). As shown in Table 5, all three groups improved over time, although more strongly for the active treatments.

**3.3.3 Procedural calculations learning—**For procedural calculations, there were no significant interactions. Considering other factors, there were main effects for pretest, $p < .0001$, for difficulty status, $p < .04$, and for treatment condition, $p < .0006$, although not for site. MD students outperformed MDRD students ($d = +0.41$). Strategic counting with deliberate practice outperformed control, $p < .0001$, $d = +0.60$, and outperformed strategic counting without deliberate practice, $p < .03$, $d = +0.21$; the other groups did not differ significantly. Table 5 shows the amount of improvement by group, which was positive for all. Special education status was also considered in supplemental analyses. No significant interactions were identified, and the treatment effect remained, $p < .0001$. However, difficulty status was no longer significant; instead, special education status was, $p < .0001$; site was now also significant, $p < .05$, with Houston students outperforming Nashville students. Among treatment groups, students who received strategic counting with deliberate practice continued to outperform the other two groups, but now students who received strategic counting without deliberate practice significantly outperformed control, $p < .01$. The reason for this pattern is that special education status was associated with weaker performance, and students in the active treatments were more likely to be identified for special education than controls.

**3.3.4 Algebraic learning—**For Find X, outcomes were analyzed dichotomously (see Preliminary Analyses). There were no interactions. There were main effects for pretest, $p < .009$, difficulty status, $p < .02$, and treatment group, $p < .0001$, although not for site, $p > .05$, controlling for pretest performance. Both active treatments outperformed control (both $p < .001$), but did not differ from one another. The proportion of students who achieved high scores in the counting strategies without and with deliberate practice condition, respectively, was 96 and 84; by contrast, the proportion in the control group was 58. Table 5 shows that improvement was larger in both active treatments.

For Number Sentences, pretest performance was considered dichotomously (as already noted), whereas posttest performance was more proportional, particularly in the active treatment groups. There were no interactions; however, there were significant effects for pretest, $p < .0001$, and treatment group, $p < .0001$, but not for site or difficulty status, controlling for all other factors. As shown in Table 5, students in the active treatment groups outperformed those in control (with practice, $p < .02$, $d = +0.55$; without practice, $p < .02$, $d = +1.03$); also, students without practice outperformed those with practice, , $p < .05$, $d = +0.51$. Although free lunch status was related to Number Sentences performance, including this as an additional covariate did not alter findings.

**3.3.5 Word-problem learning—**For Grade 2 Vanderbilt Story Problems, there were no interactions. Considering other factors, there were main effects for pretest, $p < .0001$, and for treatment condition, $p < .0004$, although not for difficulty status or site. Students in active treatments outperformed those in control (strategic counting without deliberate practice $d = +0.66$; strategic counting with deliberate practice $d = +0.51$, both $p < .001$); students in these active treatments did not differ significantly.

For Grade 3 Vanderbilt Story Problems, there was a four-way interaction of pretest, site, difficulty status, and treatment, $F(2,126) = 3.43$, $p < .04$, $\eta_p^2=.051$; however, this result should be considered with caution because it was unduly influenced by one Houston student with MD in each active treatment whose performance increased substantially from pre- to posttest; without either or both of these cases, there were no interactions of any kind. In the final model, therefore, the only significant effects were for pretest, $p < .0001$, and for treatment condition, $p < .0008$, not for difficulty status or site. Students in active treatments outperformed those in control (strategic counting without deliberate practice $d = +0.72$; strategic counting with deliberate practice $d = +0.54$, both $p < .001$); students in these active treatments did not differ significantly. Given the skewness in the posttest results, the square-root transformed variable was also evaluated; results were not substantively different (e.g., no interactions involving treatment group, with both active treatments outperforming controls), even with all observations considered. Tables 4 and 5 show the greater score improvement in the active treatments relative to the control group.

For KeyMath Problem Solving, there was a four-way interaction of pretest, site, difficulty status, and treatment condition, $F(2,126) = 3.20$, $p < .05$, $\eta_p^2=.048$; however, this result should be considered with caution because it was unduly influenced by a single control student whose performance increased substantially from pre- to posttest; without this case, there were no interactions of any kind and in the final model, the only significant effects were for pretest, $p < .0001$, and for treatment condition, $p < .03$, not for difficulty status or site. Students in active treatments outperformed those in control (strategic counting without deliberate practice $d = +0.55$; strategic counting with deliberate practice $d = +0.36$, both $p < .04$); students in these active treatments did not differ significantly. As noted earlier, subsidized lunch status and age were related to KeyMath outcomes. When models were evaluated including these additional covariates, results did not change substantively.

## 4.0 Discussion

To assess the effects of strategic counting instruction with and without deliberate practice among students with MD, we contrasted a no-tutoring control group against two variants of strategic counting instruction. Both NC remediations were embedded in the previously validated Pirate Math word-problem tutoring protocol (Fuchs et al., 2009; Fuchs et al., 2008).

In one variant, the focus on NCs was limited to a single lesson that simply taught strategic counting. In the other, students were taught counting strategies in the same single lesson but then also practiced strategic counting for answering NCs for 4–6 min in each of the 42 (of 48) intervention sessions sessions in which practice occurred (following the introductory unit). Across the study, students with deliberate practice had approximately 168–252 min of practice with the counting strategies, which the other active condition lacked. Instead, students without deliberate counting strategies practice spent that time on reading whole numbers and being provided with correct responses for the NC errors they committed.

Pinpointing the value of practice is important. Although researchers and curriculum designers assume that practice is important for students with MD (e.g., Carnine, 1997; Coyne et al., 2009; Jitendra et al., 2001), the present study was the first to isolate its effects for this population of learners. Our results suggest that practice is important. The remediation condition that included deliberate practice with the counting strategies effected superior NC fluency compared to the control condition, with a large effect size of 0.67. More importantly for the purpose of this study, students who received deliberate practice also outperformed those who were taught the counting strategies but were not provided with deliberate practice with those strategies. When comparing these two tutoring conditions, the effect size was 0.22, which meets the federal What Works Clearinghouse criterion for effective practice.

Although the effect of deliberate practice with strategic counting was clear and present, the NC outcomes of the students in the strategic counting condition with*out* deliberate practice still reliably exceeded those in the control group, with an effect size of 0.43. This is notable because this NC remediation was confined to a single lesson. After this lesson, tutors did not review the strategies; they did not demonstrate the strategies; they did not provide practice to contextualize the use of the strategies; they did not prompt students to use counting strategies to correct NC errors. So with just a single lesson in which tutors taught strategic counting, MD students realized impressive improvement in NC fluency. Although strategic counting is not typically addressed in the general education curriculum (Miller & Hudson, 2007), results suggest it should be adopted, especially for students at-risk for poor math outcomes.

The present findings compare nicely with the results of prior studies in our research program, which employed a more intensive focus on NCs. Across three previous randomized control trials, we examined the effects of drill and practice to encourage automatic retrieval, conceptual lessons to promote decomposition strategies, and (as done in the present study) the teaching of efficient counting strategies. Regardless of approach, effect sizes compared to control conditions were of similar magnitude, for all but the present study's strategic counting without deliberate practice remediation (where the effect size fell to 0.43). The generally strong findings across the four studies in this research program are probably due to the fact that the interventions shared strong instructional principles: Instruction was explicit; the instructional design attempted to minimize the learning challenge; we required practice (in all but the present study's strategic counting without deliberate practice); and we relied on motivators to help students regulate their attention and behavior.

Given similar efficacy across our approaches to NC remediation, we conclude that the strategic counting remediation with deliberate practice is most useful due to its superior efficiency. Computerized drill and practice, as with Fuchs et al. (2008), is more costly in that it requires schools to dedicate technology resources, without enjoying personnel savings. This is because computerized drill and practice requires tutor supervision, without which MD students use computers in inappropriate and inattentive ways (Fuchs, Fuchs, Hamlett, & Appleton, 2002). Conceptual instruction, as with Powell et al. (2009), is also less efficient than strategic counting with deliberate practice because its complexity requires more training and closer supervision of tutors. Future work should be conducted to estimate the costs associated with the strategic


NIH-PA Author Manuscript
NIH-PA Author Manuscript

counting remediation we investigated in the present study. Also, conclusions about the relative advantages of strategic counting, computerized drill and practice, and conceptual instruction need to qualified because these empirical comparisons are, of course, limited by the particular ways in which we implemented these approaches.

Transfer from NC remediation to procedural calculations occurred for both strategic counting conditions. In the present study, although Pirate Math word-problem tutoring directly focused on procedural calculations (with a review lesson in the introductory unit and with review/ practice as it occurred naturally within the Pirate Math lessons), the variation in the two strategic counting NC remediations contributed to differential procedural calculations outcomes. The effect size comparing the practice and no-practice conditions (0.39) was similar to the direct effect on NCs (0.43) even though both active conditions received the same dose of procedural calculations instruction. This transfer from NC remediation to procedural calculations is practically important because (a) it corroborates a previous study (Fuchs et al., 2009) where transfer to procedural calculations occurred when counting strategies are used for NC remediation (Fuchs et al.), and (b) across the four studies of our research program, such transfer occurred only when the approach to NC remediation was strategic counting. Again, this suggests the utility of strategic counting remediation. Finding evidence of transfer from NC remediation to procedural calculations is also theoretically important because NCs are viewed as a signature, bottleneck deficit for students with mathematics disability (Fleishner et al., 1982.; Geary et al., 1987; Goldman et al., 1988). With a fixed amount of attention, MD students likely allocate available resources to derive NC answers instead of focusing on the demands of the more complex mathematics into which NCs are embedded (cf. Ackerman, Anhalt, & Dykman, 1986; Goldman & Pellegrino, 1987). If NCs represent a signature deficit, performance on more complex mathematics tasks should grow as a function of increasing NC skill, 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.

By contrast, in terms of word-problem outcomes, we found no evidence to support such a hypothesis. Although students who received strategic counting with deliberate practice grew significantly better on NC fluency than students who did not receive deliberate practice, there were no reliable differences in the word-problem outcomes of these two groups of students, who all received Pirate Math word-problem tutoring. This suggests that the source of difficulty with word problems is not that students divert attention from word problems to the NCs embedded in those problems, but rather that students fail to comprehend the relations among the numbers embedded in the narratives or to process the language in those stories adequately. It also suggests that NCs are not the bottleneck for word-problem performance. Instead, difficulty with word problems implicates language (Fuchs et al., 2005, 2006). Because NC remediation transferred to procedural calculations but not to word problems, future work should continue to explore transfer from NCs to word problems as well as to other components of the mathematics curriculum.

Although there was no indication of transfer from NC remediation to word problems, results do replicate the efficacy of Pirate Math word-problem tutoring. With the schema-broadening instruction provided in Pirate Math, students in both active conditions reliably outperformed control students on each word-problem outcome, with no differences between the two active conditions on word problems. Effects favoring word-problem tutoring were more consistent when word-problem measures were better aligned with tutoring, even though none of the problems on any outcome measures had been used for instruction. Both versions of Vanderbilt Story Problems, one assessing the second-grade curriculum and the other assessing the third-grade curriculum, were limited to the types of problems directly addressed in Pirate Math (total, difference, and change relationships) with missing information in all three positions, with and without irrelevant information, and with and without charts/graphs. On these measures, Pirate Matheffected superior outcomes compared to the control group, with large effect sizes for both active conditions. On KeyMath, which was a more distal measure because it assessed taught as well as untaught problem types, Pirate Math students again outperformed the control group, this time with more moderate effect sizes. This supports the need to align word-problem instruction with the desired outcomes by addressing the complete set of problem types assessed on those measures.

Algebra is an aspect of mathematics conventionally taught much later in the curriculum. With Pirate Math, students are taught to solve for an unknown variable in simple addition and subtraction equations and, more importantly with respect to algebraic reasoning, to represent word-problem schema using algebraic equations. On Find X, where students solve algebraic equations, both Pirate Math conditions outperformed control. The proportion of students who achieved high scores in the strategic counting with and without deliberate practice condition, respectively, was 84 and 96; the proportion in the control group was smaller (58). On Number Sentences, where students generate algebraic equations to represent word-problem models without solving those equations, both active treatments again performed better than the control group with large effect sizes of 0.55 and 1.03. These findings suggest that algebraic cognition improves when algebra is taught as a tool for solving word problems. This is notable because study participants were young and were severely deficient in incoming math skill. 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 earlier in the curriculum may represent a productive innovation. It is important to note, however, that in terms of the effects of practice, this was the one measure where the strategic counting without deliberate practice condition outperformed the strategic counting with deliberate practice condition, thereby raising the possibility that practice on NCs somehow undermined students' algebraic cognition. Future research should investigate this possibility.

We note two limitations to our study of practice effects. First, although we substantiated the importance of practice, results do not provide the basis for determining what amount of practice is optimal. Additional research is needed to provide such a recommendation. Second, readers should note that several differences existed at the beginning of the study among the treatment conditions. This concern is, however, mitigated by the fact that all of these differences favored the control group over both active tutoring conditions. That is, among MDRD (but not MD) students, the control group began the study performing higher than both groups of tutored students on Grade 2 Story Problems; across MDRD and MD students, performance on the Iowa favored the control group over both groups of tutored students; and a higher proportion of students qualified for special education in both tutored groups compared to the control group. All of these differences create a stringent test of tutoring efficacy. In fact, the control group completed the study no longer enjoying any advantage on any measure and suffering reliably lower performance on all measures except the Iowa (where their pretreatment advantage has been erased).

Besides assessing the efficacy of the tutoring protocols, a second purpose of the present study was to examine whether remediation effects depend 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 an interaction whereby MDRD students, but not MD students, would require deliberate practice with the counting strategies. In addition, because using text to construct a word-problem model involves language (e.g., Fuchs et al., 2005; Fuchs et al., 2006; Fuchs, Fuchs, Stuebing et al., 2008; Swanson & Beebe-Frankenberger, 2004; Swanson, 2006) and because the language of students with MDRD is depressed

NIH-PA Author Manuscript

NIH-PA Author Manuscript

compared to students with MD (Powell et al., 2009), we hypothesized that MDRD students would be less responsive to Pirate Math word-problem tutoring. We found no evidence of differential responsiveness to intervention as a function of difficulty status on any outcome. This raises questions about the tenability of the MD/MDRD subtyping scheme and suggests the need to pursue other subtyping frameworks (Fletcher et al., 2007). For example, some work (Fuchs, Fuchs, Stuebing et al., 2008) suggests that calculations disability versus word-problem disability may represent a more productive subtyping scheme. We note, however, that despite the lack of differential responsiveness to treatment, students with MD (across tutoring conditions and sites) did outperform students with MDRD on some math skills (i.e., WRAT-Arithmetic, Number Sentences, and KeyMath at pretest; NCs, procedural calculations, and Find X at posttest). Therefore, additional work is warranted to examine the tenability of the MD/MDRD subtyping scheme, even as research pursuing alternative frameworks proceeds.

## Acknowledgments

The research described in this paper was supported in part by Core Grant Number P30HD15052 from the National Institute of Child Health and Human Development to Vanderbilt University and by Award Number P01046261 from the National Institute of Child Health and Human Development 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 National Institute of Child Health and Human Development 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 sequaelae. 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.

Baroody AJ, Bajwa NP, Eiland M. Why can't Johnny remember the basic facts? Developmental Disabilities Research Reviews 2009;15:69–79. [PubMed: 19213010]

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.

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.

Carnine D. Instructional design in mathematics for students with learning disabilities. Journal of Learning Disabilities 1997;30:130–141. [PubMed: 9066274]

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.

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

Coyne M, Zipoli R, Chard D, Faggella-Luby M, Ruby M, Santoro L, Baker S. Reading and Writing Quarterly 2009;25:221–245.

Fleishner 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.

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, Compton DL, Powell SR, Seethaler PM, Capizzi AM, Schatschneider C, Fletcher JM. The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. Journal of Educational Psychology 2006;98:29–43.

Fuchs LS, Fuchs D, Hamlett CL, Appleton AC. Explicitly teaching for transfer: Effects on the mathematical problem solving performance of students with disabilities. Learning Disabilities Research and Practice 2002;17:90–106.

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. [PubMed: 20057912]

Fuchs, LS.; Powell, SR.; Hamlett, CL.; Powell, SR. Third-grade math battery. Available from L.S. Fuchs, 228 Peabody, Vanderbilt University; Nashville, TN 37203: 2003.

Fuchs LS, Powell SR, Hamlett CL, Fuchs D, Cirino PT, Fletcher JM. Remediating computational deficits at third grade: A randomized field trial. Journal of Research on Educational Effectiveness 2008;1:2–32.

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

Fuchs, LS.; Seethaler, PM. Find X, Number Sentences, and Vanderbilt Story Problems. Available from L.S. Fuchs, 228 Peabody, Vanderbilt University; Nashville, TN 37203: 2008.

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

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, Hoard MK, Byrd-Craven J, Nugent L, Numtee C. Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. Child Development 2007;78:1343–1459. [PubMed: 17650142]

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, Hamson CO. Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. Journal of Experimental Child Psychology 1999;74:213–239. [PubMed: 10527555]

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.

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.

Groen GJ, Parkman JM. A chronometric analysis of simple addition. Psychological Review 1972;79:329–343.

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, Nolet V, Ping Xin Y, Gomez O, Renouf K, Iskold L, DaCosta J. An analysis of middle school geography textbooks: Implications for students with learning problems. Reading & Writing Quarterly: Overcoming Learning Difficulties 2001;17:151–173.

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, 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, 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.

Kroesbergen EH, Van Luit JEH, Maas CJM. Effectiveness of explicit and constructivist instruction for low-achieving students in The Netherlands. Elementary School Journal 2004;104:233–251.

Mazzocco MMM. An introduction to the special issue: Pathways to mathematics learning difficulties and disabilities. Developmental Disabilities Research and Reviews 2007;15(1):1–3.

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. Effects of fact retieval tutoring on third-grade students with math difficulties with and without reading difficulties. Learning Disabilities Research and Practice 2009;24:1–11. [PubMed: 19448840]

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.

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

Shalev RS, Auerbach J, Manor O, Gross-Tsur V. Developmental dyscalculia: Prevalence and prognosis. European Child and Adolescent Psychiatry 2000;9(2):58–64.

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, Inc; 1993.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

**Table 1**

Demographic and Screening Data by Tutoring Condition and by Difficulty Status

| | Tutoring Condition | | | Difficulty Status | |
|---|---|---|---|---|---|
| | Strategic Counting with Practice | Strategic Counting without Practice | Control | MD | MDRD |
| **Variable** | **(n=51)** | **(n=49)** | **(n=50)** | **(n=41)** | **(n=109)** |
| Age in Years | 8.31 (0.45) | 8.38 (0.51) | 8.41 (0.58) | 8.05 (0.32)[a] | 8.49 (0.53)[b] |
| Female | 39% | 43% | 48% | 44% | 43% |
| Subsidized Lunch | 73% | 71% | 82% | 61% | 81% |
| Special Education | 43% | 45% | 18% | 17% | 42% |
| Retained | 31% | 35% | 32% | 12% | 40% |
| English Second Lang. | 18% | 16% | 24% | 17% | 20% |
| African American | 59% | 63% | 46% | 51% | 58% |
| Caucasian | 8% | 14% | 10% | 12% | 10% |
| Hispanic | 31% | 20% | 36% | 37% | 27% |
| Other | 2% | 2% | 8% | 0% | 5% |
| WASI FSIQ | 84.78 (10.3) | 87.04 (12.7) | 85.24 (11.2) | 91.15 (12.3)[a] | 83.61 (10.4)[b] |
| Iowa Problem Solving | 22.47 (8.4) | 23.24 (9.8) | 26.32 (8.6) | 25.05 (7.0) | 23.61 (9.7) |
| WRAT Arithmetic | 80.45 (8.3) | 81.82 (7.6) | 81.00 (7.2) | 84.93 (5.5)[a] | 79.63 (7.9)[b] |
| WRAT Reading | 84.57 (15.7) | 85.94 (14.2) | 85.20 (12.8) | 103.49 (6.5)[a] | 78.36 (9.4)[a] |

Note. Percentages are computed relative to the number of individuals in that group (e.g., 39% of the 51 students in strategic counting with practice were female). Percentages for ethnicity within a column total of 100%. WASI=Wechsler Abbreviated Scales of Intelligence; WRAT=Wide Range Achievement Test-3$^{rd}$ ed. Iowa Problem Solving score is a normal curve equivalent. Numbers in parentheses are *SD*s. For mean scores, values in a given row with different superscripts are significantly different from one another.

## Table 2

Pretest Performance (Standardized Means, SD) by Tutoring Condition and Difficulty Status

| | Tutoring Condition | | | | Difficulty Status | | |
|---|---|---|---|---|---|---|---|
| | | Strategic Counting with Practice | Strategic Counting without Practice | Control | | MD | MDRD |
| **Variable** | ***F*** | **(n=51)** | **(n=49)** | **(n=50)** | ***F*** | **(n=41)** | **(n=109)** |
| WRAT Arithmetic | (2,145),< 1 | −0.08 (1.07) | +0.09 (1.00) | −0.01 (0.94) | (1,145), 15.08** | +0.50 (0.71) | −0.19 (1.03) |
| Iowa | (2,145), 3.28* | −0.17 (0.93) | −0.08(1.09) | 0.26 (0.95) | (1,145), 1.78 | +0.12 (0.78) | −0.04 (1.07) |
| NCs | (2,136), < 1 | −0.06 (0.92) | −0.01 (0.80) | +0.07 (1.00) | (1,136), <1 | +0.01 (1.05) | −0.00 (0.85) |
| Procedural Calc. | (2,145), < 1 | −0.13 (1.08) | +0.14 (1.01) | −0.00 (0.91) | (1,145), 2.80 | +0.16 (1.16) | −0.06 (0.93) |
| Find X | (2, N=150), 6.73* | 39% | 63% | 52% | (1, N=150), < 1 | 51% | 51% |
| Number Sentences | (2, N=150), 1.19, < 1 | 69% | 73% | 60% | (1, N=150), 1.30 | 71% | 66% |
| Grade 2 VSP | (2,138), 2.02 | −0.25 (0.95) | +0.21 (1.01) | +0.06 (1.01) | (1,138), 2.70 | +0.17 (1.09) | −0.07 (0.96) |
| Grade 3 VSP | (2,145), 1.92 | −0.20 (0.88) | +0.20 (1.13) | +0.01 (0.96) | (1,145), 3.59 | +0.20 (0.98) | −0.08 (1.00) |
| KeyMath | (2,145), 2.59 | −0.19 (1.03) | +0.26 (0.98) | −0.06 (0.96) | (1,145), 11.11** | +0.43 (0.98) | −0.16 (0.96) |

Note. VSP is Vanderbilt Story Problems. *F* values are for the main effects, controlling for other factors in model; for Grade 2 VSP, there were there were interactions of tutoring group and/or difficulty status with one another or with site. Numbers in parentheses are *df*s. Scores are *z*-standardized raw means (uncorrected for all model factors). NCs is a factor scores derived from four measures. Values for Find X and for Number Sentences are Wald chi-square statistics based on the proportion of students scoring high (4–8 for Find X and 8–16 for Number Sentences).

* $p < .05$;

** $p < .01$; others are not significant.

**Table 3**

F Values at Posttest[a]

| | Effects | | | |
|---|---|---|---|---|
| **Outcome Measures** | **Pretest (1)** | **Difficulty Status (1)** | **Treatment (2)** | **Site (1)** |
| NCs (142) | 104.22*** | 15.08*** | 12.87*** | < 1 |
| Procedural Calc. (144) | 52.89*** | 4.45* | 7.77*** | 2.54 |
| Find X (150) | 6.89** | 5.52* | 18.27*** | < 1 |
| Number Sentences (128) | 23.97*** | 2.08 | 5.21** | < 1 |
| Grade 2 VSP (142) | 47.73*** | < 1 | 8.34*** | < 1 |
| Grade 3 V SP (142) | 25.35*** | 1.42 | 7.57*** | < 1 |
| KeyMath (143) | 64.45*** | 2.06 | 3.65* | < 1 |

[a]Denominator *dfs* are in parentheses next to the measures; numerator *dfs* are next to the effects, which change depending on the number of terms in the final model. VSP is Vanderbilt Story Problems. For procedural calculations, results are without special education status. For Find X, overall N is in parenthesis in the first column; effects are Wald chi-square values. For Grade 3 VSP, results are without the two outliers. For KeyMath, results are without outlier and without free lunch status and age covariates.

* $p$<.05;

** $p$<.01;

*** $p$<.001; others are not significant.

## Table 4

Posttest Performance (Standardized Means, SD) by Tutoring Condition and Difficulty Status

| | Tutoring Condition | | | Difficulty Status | |
|---|---|---|---|---|---|
| | **Strategic Counting with Practice** | **Strategic Counting without Practice** | **Control** | **MD** | **MDRD** |
| **Variable** | **(n=51)** | **(n=49)** | **(n=50)** | **(n=41)** | **(n=109)** |
| NCs | +0.27 (0.88) | +0.07 (0.95) | −0.34 (0.94) | +0.37 (1.06) | −0.14 (0.87) |
| Procedural Calculations | +0.26 (0.95) | +0.06 (1.02) | −0.33 (1.01) | +0.30 (0.97) | −0.11 (0.99) |
| Find X | 84% | 96% | 58% | 93% | 75% |
| Number Sentences | +0.01 (0.90) | +0.49 (0.98) | −0.49 (0.89) | +0.24 (1.19) | −0.09 (0.91) |
| Grade 2 VSP | +0.11 (0.92) | +0.30 (1.06) | −0.35 (0.89) | +0.22 (1.01) | −0.05 (0.98) |
| Grade 3 VSP | +0.04 (0.85) | +0.28 (1.08) | −0.41 (0.82) | +0.17 (0.90) | −0.11 (0.97) |
| KeyMath | +0.02 (0.87) | +0.23 (1.02) | −0.32 (0.96) | +0.39 (0.80) | −0.17 (0.99) |

Note. VSP is Vanderbilt Story Problems. Scores are *z*-standardized raw means (uncorrected for all model factors), except for Find X, where values are the proportion of students scoring high (4–8). Numbers are from final models in Table 3.

## Table 5

Pretest and Posttest Performance (Raw Means, SDs) by Tutoring Condition

| | Tutoring Condition | | | | | |
|---|---|---|---|---|---|---|
| | Strategic Counting with Practice | | Strategic Counting without Practice | | Control | |
| Variable | (n=51) | | (n=49) | | (n=50) | |
| | Pre | Post | Pre | Post | Pre | Post |
| NCs (100) | 28.18 (17.2) | 48.82 (15.8) | 29.69(13.7) | 45.29 (17.0) | 28.92(17.2) | 37.8 (16.9) |
| Procedural Calculations (25) | 6.02 (4.9) | 12.49 (5.4) | 7.27 (4.6) | 11.35 (5.5) | 6.60 (4.1) | 9.14 (5.7) |
| Find X (8) | 2.57 (2.8) | 5.37 (2.6) | 3.96 (2.8) | 6.73 (1.9) | 3.46 (3.2) | 3.94 (3.42) |
| Number Sentences (8) | 1.59 (1.4) | 3.16 (1.9) | 1.96 (1.6) | 4.18 (2.1) | 1.56(1.6), | 2.10 (1.9) |
| Grade 2 VSP (30) | 7.10 (4.5) | 16.11 (7.1) | 9.31(4.8) | 17.58 (8.2) | 8.57 (4.8) | 12.55 (6.8) |
| Grade 3 VSP (30) | 3.92 (2.7) | 10.08 (5.5) | 5.35 (3.8) | 11.58 (6.9) | 4.74(3.2) | 7.18 (5.2) |
| KeyMath (12) | 2.33 (1.5) | 4.20 (2.1) | 3.32 (1.7) | 4.94 (2.4) | 2.65 (1.4) | 3.45 (2.0) |

Note. VSP is Vanderbilt Story Problems. As with Table 4, numbers are from final models in Table 3. All other data and statistics refer to standardized scores.

## Table 6

Effect Sizes by Tutoring Condition and Difficulty Status

| | Contrast Tutoring Condition | | | Contrast Difficulty Status |
|---|---|---|---|---|
| | Strategic Counting with Practice v. | | Strategic Counting without Practice v. | MD v. |
| Variable | Control | without Practice | Control | MDRD |
| NCs | +0.67 (+0.26 to +1.06) | +0.22 (−0.17 to +0.62) | +0.43 (+0.02 to +0.83) | +0.55 (+0.19 to +0.92) |
| Procedural Calc. | +0.60 (+0.20 to +1.00) | +0.21 (−0.18 to +0.60) | +0.39 (−0.01 to +0.79) | +0.41 (+0.05 to +0.77) |
| Number Sentences | +0.55 (+0.15 to +0.94) | −0.51 (−0.90 to −0.11) | +1.03 (+0.61 to +1.45) | +0.33 (−0.03 to +0.70) |
| Grade 2 VSP | +0.51 (+0.11 to +0.90) | −0.19 (−0.59 to +0.20) | +0.66 (+0.26 to +1.07) | +0.27 (−0.09 to +0.64) |
| Grade 3 VSP | +0.54 (+0.14 to +0.94) | −0.24 (−0.64 to +0.16) | +0.72 (+0.31 to +1.13) | +0.29 (−0.09 to +0.64) |
| KeyMath | +0.36 (−0.03 to +0.76) | −0.22 (−0.61 to +0.17) | +0.55 (+0.15 to +0.95) | +0.59 (+0.23 to +0.96) |

Note. Effect sizes are based unadjusted means at posttest and pooled standard deviation of the two groups compared. Effect sizes for diagnosis (MD v. MDRD) are based on the means that appear in Table 4, but may appear larger than what might be inferred from Table 3 because significance (although not actual means) in Table 3 is based on all factors in the model. Positive effect sizes indicate higher means for the first group named in the comparison. Find X is not in this table because results were dichotomized; here, 47/49 with practice, 43/51 without practice, and 29/50 controls achieved high scores (4–8). Values in parentheses are confidence intervals for the effect sizes.

NIH-PA Author Manuscript
NIH-PA Author Manuscript



## Appendix A

Sample Problems by Problem Type And by Position of Missing Information (A+B=C or A−B=C)[a]

| 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=C) |
| 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) |