Published in final edited form as:

J Res Educ Eff. 2008 ; 1(1): 2–32. doi:10.1080/19345740701692449.

Remediating Computational Deficits at Third Grade: A Randomized Field Trial

Lynn S. Fuchs, Sarah R. Powell, Carol L. Hamlett, and Douglas Fuchs

Department of Special Education, Vanderbilt University, Nashville, Tennessee, USA

Paul T. Cirino and Jack M. Fletcher

Department of Psychology, University of Houston, Houston, Texas, USA

Abstract

The major purposes of this study were to assess the efficacy of tutoring to remediate 3rd-grade computational deficits and to explore whether remediation is differentially efficacious depending on whether students experience mathematics difficulty alone or concomitantly with reading difficulty. At 2 sites, 127 students were stratified on mathematics difficulty status and randomly assigned to 4 conditions: word recognition (control) tutoring or 1 of 3 computation tutoring conditions: fact retrieval, procedural computation and computational estimation, and combined (fact retrieval + procedural computation and computational estimation). Results revealed that fact retrieval tutoring enhanced fact retrieval skill, and procedural computation and computational estimation tutoring (whether in isolation or combined with fact retrieval tutoring) enhanced computational estimation skill. Remediation was not differentially efficacious as a function of students' mathematics difficulty status.

Keywords

Mathematics difficulty; remediation; computation

Approximately 4 to 1% of the school-age population suffers from mathematics disability (e.g., Badian, 1983; Shalev, Auerbach, Manor, & Gross-Tsur, 2000). Although the prevalence of mathematics disability is similar to that of reading disability, less systematic study has been directed at mathematics disability (Rasanen & Ahonen, 1995) despite evidence that poor mathematics skills are associated with lifelong difficulties in school and in the workplace. Mathematics competence, for example, accounts for variance in employment, income, and work productivity even after intelligence and reading have been explained (Rivera-Batiz, 1992).

Some research illustrates how prevention activities at preschool (e.g., Clements & Sarama, 2007), kindergarten (e.g., Griffin, Case, & Siegler, 1994), or first grade (e.g., Fuchs, Fuchs, Yazdian, & Powell, 2002) can substantially improve math performance. For example, at the beginning of first grade, Fuchs et al. (2005) identified 169 students in 41 classrooms as at risk for math difficulties based on their low initial performance. These children were randomly assigned to a control group or to receive small-group tutoring that occurred three times per week for 20 weeks. Results showed that math development across first grade was significantly and substantially superior for the tutored group than for the control group on

computation, concepts, applications, and story problems. (We note that the math literature relies on the terms *story problems* and *word problems* interchangeably; in this article, we use the term *story problems*.) In addition, the incidence of students with mathematics disability was substantially reduced at the end of first grade, and this reduction in mathematics disability remained in the spring of second grade, 1 year after tutoring ended (Compton, Fuchs, & Fuchs, 2006). Nevertheless, despite the efficacy of tutoring, students were not universally responsive. A subset of the tutored students manifested severe mathematics deficits: approximately 3 to 6% of the school population (depending on the measure and cut-point for the severity of mathematics performance).

As this example illustrates, the need for remediation among some students will persist even when prevention services are generally effective. Unfortunately, the literature on the remediation of mathematics deficits is limited in important ways. This research article describes a randomized controlled trial exploring the efficacy of remedial tutoring for students identified with low math performance at the beginning of third grade. The remedial tutoring was designed to enhance performance on three aspects of computational skill in addition and subtraction: (a) number combinations, where problems involve single-digit operands and can be solved via counting or committed to long-term memory for automatic retrieval; (b) exact procedural skill on problems with two-digit operands; and (c) computational estimation of problems with two-digit operands. Some limited, prior remediation work exists on number combinations and exact procedural skill. In this introduction, we summarize that body of work. Then we discuss the possibility that remediation efficacy may be affected by whether students experience mathematics difficulty alone or in combination with reading problems. Finally, we clarify the purposes of our study.

PRIOR REMEDIATION WORK

Number combinations are an important component of the initial mathematics curriculum. Research shows that number combination skill is a significant path to performance in procedural computation as well as in word problems (Fuchs, Fuchs, Compton, et al., 2006). Conventionally, number combinations are incorporated into the curriculum at kindergarten, first grade, and second grade, and typically developing students are on their way toward automatic retrieval of number combinations by the beginning of third grade. Therefore, when students still manifest deficiencies involving number combinations at the beginning of third grade, a pressing need exists for remediation.

Unfortunately, prior work examining remediation efficacy has not relied on well-designed randomized control studies. Okolo (1992) and Christensen and Gerber (1990) both contrasted computer-assisted instruction on number combinations in a gamelike format to computer-assisted instruction in an unadorned drill format. For students with learning disabilities at Grades 3 to 6, Okolo found no significant differences between groups, both of which significantly improved their number combinations proficiency. At Grades 4 to 6, Christensen and Gerber found that students with learning disabilities were disadvantaged by a gamelike format, perhaps due to the distracting nature of the presentation. Neither study, however, incorporated a control group to assess whether the computer-assisted instruction condition effected better outcomes than would have been expected. More recently, Tournaki (2003) questioned the value of paper-pencil drill and practice by contrasting this approach with instruction designed to teach students to count answers to arithmetic problems in a strategic fashion. With 8- to 10-year-old students with learning disabilities, results revealed an advantage for strategic counting over rote practice, but this result was not surprising because the format of that practice condition did not incorporate well-established instructional design features for promoting learning. Instead, the paper-pencil practice

provided feedback on a delayed schedule, without deliberately mixing known with unknown combinations and without systematic review of mastered combinations; by contrast, the strategy instruction incorporated immediate corrective feedback and reteaching whenever an error occurred.

Consequently, a need exists for well-designed randomized controlled studies that test methods for remediating students' deficits with number combinations to promote automatic fact retrieval. At the same time, the literature provides less guidance about remediating procedural skill deficits. Fuchs, Fuchs, Hamlett, and Stecker (1991) designed an expert system to incorporate pictorial representations, procedures for supplying clear explanations, and verbal rehearsal with fading for teaching procedural math, including multidigit computation. This system was tested experimentally with 33 teachers who provided math instruction to students with learning disabilities in Grades 2 to 8. Teachers were assigned randomly to three conditions: (a) ongoing, systematic assessment of student growth with descriptive profiles of students' strengths/weaknesses; (b) ongoing, systematic assessment of student growth with descriptive profiles of students' strengths/weaknesses plus use of the expert system; or (c) typical practice controls. Teachers implemented tutoring for 20 weeks. Analyses indicated that only the expert system group effected superior learning relative to the controls. This study was, however, conducted with students across a wide grade span, and effects were not disaggregated by grade or by type of computational skill. Nonetheless, these findings provide evidence that the combination of these instructional principles may be efficacious in remediating procedural computational skill deficits.

MATHEMATICS DIFFICULTY ALONE OR IN COMBINATION WITH READING DIFFICULTY

The relation between reading and mathematics performance is moderate to high (Aiken, 1972; Wilkinson, 1993). Lewis, Hitch, and Walker (1994) estimated that mathematics deficits co-occur in approximately 40% of individuals with reading disability (also see Delazer & Bartha, 2001; Shaywitz et al., 1999). Moreover, research indicates that reading skill may influence relations between cognitive characteristics that affect math competence (Fuchs, Fuchs, Compton, et al., 2006). As might be expected, therefore, some previous work suggests that students with concurrent difficulty in math and reading experience more pervasive difficulties, including more severe mathematics deficits in procedural computation (e.g., Jordan & Hanich, 2000), simple word problems (e.g., Hanich, Jordan, Kaplan, & Dick, 2001; Jordan & Hanich, 2000), and complex forms of mathematical problem solving (Fuchs & Fuchs, 2002). By contrast, other work (e.g., Jordan, Hanich, & Kaplan, 2003; Landerl, Began, & Butterworth, 2004) shows that students with mathematics disabilities, with and without concurrent reading difficulty, experience comparable mathematics deficits.

Most recently, Cirino, Ewing-Cobbs, Barnes, Fuchs, and Fletcher (2007) assessed 291 third and fourth graders whom the researchers classified as having math disabilities, reading disabilities, math and reading disabilities, or normal achievement. Students were assessed on computerized measures of cognitive arithmetic, addition, subtraction, and estimation. Problems were categorized as having small operands (addends of 2 to 5 with answers between 5 and 9; subtrahends of 2 to 6 with answers between 2 and 7) or large operands (addends of 3 to 9 with answers between 11 and 17; subtrahends of 4 to 9 with answers between 4 and 15; some problems required regrouping). Students with mathematics disability differed from those with mathematics and reading disability only on accuracy for small operands subtraction problems.

As illustrated in a review of this evidence by Fletcher, Lyon, Fuchs, and Barnes (2007), research findings are inconsistent about whether groups of students specifically with

mathematics difficulty differ on various aspects of mathematics skill from students with concurrent mathematics and reading difficulty. Most studies addressing this question have employed a cross-sectional causal comparative design, whereby subgroups are identified that generally meet low performance criteria in mathematics or in mathematics and reading; then these subgroups are assessed on measures that tap various aspects of mathematical cognition. An alternative approach for addressing the same issue is experimental, whereby students who meet classification criteria for the two subtypes of mathematics difficulty are randomly assigned to receive targeted instruction in specific mathematical processes or to participate in a control condition. The goal is to determine whether the subgroups of learners respond differentially to the instruction. This design offers the basis for stronger causal inferences.

We identified only one study adopting this alternative approach. Fuchs, Fuchs, and Prentice (2004) randomly assigned third-grade classrooms to validated math problem-solving instruction (experimental) or to teacher-designed instruction (control), and the validated problem-solving instruction was delivered in a whole-class format over 16 weeks with strong fidelity. The researchers retrospectively identified students, based on preceding spring TerraNova scores (CTB/McGraw-Hill, 2003), who manifested difficulty in mathematics and reading, difficulty with mathematics alone, difficulty with reading alone, and without either type of difficulty. Effects of the 16-week intervention were then assessed on problem-solving measures using score (problem solving, computation, communication) as a within-subjects variable; the between-subject variables were tutoring condition and type of difficulty. On computation and communication scores, students with any of the three forms of difficulty were differentially less responsive than students without difficulty. By contrast, on the problem-solving score, only students with concurrent difficulty in mathematics and reading were differentially less responsive than students without difficulty. Although these findings suggest the viability of the two subtypes of mathematics difficulty, the study was conducted retrospectively, with relatively small numbers of students identified with each form of difficulty. Moreover, it addressed responsiveness to instruction only on mathematical problem solving, which may represent an aspect of mathematical cognition that is distinct from computational skill (Fuchs, Fuchs, Stuebing, et al., in press).

THE PRESENT STUDY

In the present study, we had three primary purposes. The first was to assess the efficacy of tutoring procedures for remediating mathematics deficits that had accrued by the beginning of third grade. The second purpose was to explore whether remediation was differentially efficacious depending on students' difficulty status: mathematics difficulty alone versus mathematics difficulty with concomitant reading difficulty. Finally, by running the study in two sites, one of which was distal to the development site, we asked whether the training methods were transportable and replicable. Toward these ends, we conducted a randomized field trial, whereby students were stratified on difficulty status and then randomly assigned to different tutoring conditions. One tutoring condition targeted automatic retrieval of number combinations (referred to as *fact retrieval* in this article); the second tutoring condition targeted procedural computation and computational estimation; the third tutoring condition had a combined focus on fact retrieval and procedural computation/computational estimation; and to control for attention and instructional time, the fourth (control) condition provided similar amounts of tutoring but focused on word recognition reading skill. In addition, concerned about the portability of the tutoring procedures, we developed standardized protocols that relied in part on computer-guided instruction. To assess transferability of the tutoring protocols, we conducted the study at two sites, one local to the intervention developers and a second at a distal site, and examined fidelity and tutoring

effects as a function of site. Finally, we included a variety of measures that allowed us to assess the specificity of effects.

METHOD

Participants

Across 18 schools, in 56 third-grade classrooms in Nashville and 24 third-grade classrooms in Houston, we administered the Wide Range Achievement Test 3 (WRAT–Arithmetic (Wilkinson, 1993) in large-group format to all students on whom we had obtained parental consent ($n = 1,163$). Three hundred ten (28%) met the WRAT–Arithmetic screening criterion of scoring below the 26th percentile. Of these 310, 58 were not screened further to distribute potential tutoring resources across schools and classrooms. The remaining students were assessed individually on WRAT–Reading and on the two-subtest version of the Weschler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1999). All students scoring between the 26th and 39th percentile on WRAT–Reading ($n = 45$) or earning a T score below 36 on both WASI subtests (Vocabulary and Matrix Reasoning; $n = 4$) were excluded. We identified students as having math difficulty alone (MD; WRAT–Arithmetic < 26th percentile, but WRAT–Reading > 39th percentile) or as having math and reading difficulty (MDRD; WRAT–Arithmetic < 26th percentile and WRAT–Reading < 26th percentile).

Thirty-eight students moved after screening and before randomization. The remaining 165 students were randomized to one of the four tutoring conditions. One hundred thirty-two students began intervention; the 33 who did not were primarily from Houston ($n = 31$) where screening was conducted earlier than in Nashville and where the student mobility rate is higher. The reasons these 33 students did not begin intervention were (a) moving out of the school ($n = 24$), (b) exclusionary criterion not known at the time of randomization (medical or psychiatric illness, $n = 3$), and (c) school or project scheduling conflicts (e.g., inability to fit the intervention into the school schedule, $n = 6$). Students who did not begin intervention did not differ from those who did in terms of WRAT performance, age, and ethnicity ($p > .05$). One additional student assigned to intervention mistakenly received the incorrect intervention; results did not change by excluding this student, so we retained this student in the analysis. Finally, 4 students moved during the school year, such that 128 completed intervention.

Random assignment to the four tutoring conditions occurred by blocking on site (Nashville vs. Houston) and on difficulty status (MD vs. MDRD). The four tutoring conditions were fact retrieval tutoring, procedural/estimation tutoring, fact retrieval + procedural/estimation (i.e., combined) tutoring, and word-identification tutoring. The demographics of the sample (English as Second Language status, subsidized lunch status, special education status, ethnicity, sex) did not differ as a function of site, tutoring condition, difficulty status, or site within tutoring condition. In addition, sites did not differ in the proportion of students across the difficulty status, tutoring condition, or difficulty status by tutoring conditions. Age was only weakly related to performance on outcome measures at pre- and posttest occasions and therefore was not included in analyses. Demographic and screening data are presented in Table 1 by tutoring condition and by difficulty status; pretest performance data are presented in Table 2 by tutoring condition and by difficulty status (as expected, on IQ and WRAT–Arithmetic, MD > MDRD; this would have been the case also for WRAT–Reading except the we controlled for reading in the analyses). In no case did including site as a factor change the interpretation of results on the primary outcome measures.¹ Therefore, and for ease of interpretation, site was not included in the tables, and it was trimmed as a factor from the analyses reported next.

Measures

Specific tests were used for screening (WRAT–Arithmetic, WRAT–Reading, and WASI) and for evaluating tutoring effects at pre- and posttreatment. Some measures were directly aligned with tutoring conditions. We considered these tutoring-aligned measures indexes of transfer for the tutoring conditions with which they were not aligned. In addition, other measures were not aligned with any tutoring condition and therefore represented measures of transfer across conditions.

Screening—With WRAT–Arithmetic (Wilkinson, 1993), students have 10 min to complete calculation problems of increasing difficulty. If the basal is not met, students read numerals aloud. Median reliability is .94 for ages 5–12 years. With WRAT–Reading (Wilkinson, 1993), students read aloud letters and words until a ceiling is reached. Reliability is .94. The WASI (Psychological Corporation, 1999) measures cognitive functioning with two subtests. *Vocabulary* assesses expressive vocabulary, verbal knowledge, memory, and learning ability, as well as crystallized and general intelligence with four pictures and 37 words. Participants name pictures and define words. *Matrix Reasoning* measures nonverbal fluid reasoning and general intelligence with 35 items. Examinees select an option (five choices) that best completes a visual pattern. Subtest scores are combined to yield an Estimated Full Scale IQ score, which correlates with the Wechsler Intelligence Scale for Children–III IQ score at .82. The standardization sample is 1,100 children (ages 6–16). Reliability exceeds .92.

Outcome Measures Aligned With Fact Retrieval Tutoring—The Grade 3 Math Battery (Fuchs, Hamlett, & Powell, 2003) incorporates two fact retrieval subtests. *Addition Fact Fluency 0–8* comprises 25 addition fact problems with sums from 0 to 8, presented vertically on one page. Students have 1 min to write answers. The score is the number of correct answers. Percentage of agreement, calculated on 20% of protocols by two independent scorers, was 99.9. Coefficient alpha on this sample was .93. *Subtraction Fact Fluency 0–8* comprises 25 subtraction fact problems with minuends from 0 to 8, presented vertically on one page. Students have 1 min to write answers. The score is the number of correct answers. Percentage of agreement, calculated on 20% of protocols by two independent scorers, was 98.1. Coefficient alpha on this sample was .90. *Addition Fact Fluency 0–12* comprises 25 addition fact problems with sums from 0 to 12, presented vertically on one page. Students have 1 min to write answers. The score is the number of correct answers. Percentage of agreement, calculated on 20% of protocols by two independent scorers, was 98.9. Coefficient alpha on this sample was .89. *Subtraction Fact Fluency 0–12* comprises 25 subtraction fact problems with minuends from 0 to 12, presented vertically on one page. Students have 1 min to write answers. The score is the number of correct answers. Percentage of agreement, calculated on 20% of protocols by two independent scorers, was 98.3. Coefficient alpha on this sample was .92. *Addition Fact Fluency 0–18* comprises 25 addition fact problems with sums from 0 to 18, presented vertically on one page. Students have 1 min to write answers. The score is the number of correct answers. Percentage of agreement, calculated on 20% of protocols by two independent scorers, was 99.9. Coefficient alpha on this sample was .88. *Subtraction Fact Fluency 0–18* comprises 25 subtraction fact problems with minuends from 0 to 18, presented vertically on one page. Students have 1 min to write answers. The score is the number of correct answers. Percentage of agreement, calculated on 20% of protocols by two independent scorers, was 99.3. Coefficient alpha on this sample was .89.

¹For story problems, the three-way interaction was significant. Houston MDRD controls scored better than Nashville MDRD controls, which does not alter the fact that tutoring condition did not enhance story problem performance.

Outcome Measures Aligned With Procedural/Estimation Skill Tutoring—The Grade 3 Math Battery (Fuchs et al., 2003) also includes two subtests of procedural computation. *Double-Digit Addition* provides students with 3 min to complete 20 two-digit by two-digit addition problems with and without regrouping. The score is the number of correct answers. Percentage of agreement, calculated on 20% of protocols by two independent scorers, was 99.3. In the study presented here, coefficient alpha was .91. On a larger representative third-grade sample, criterion validity with the previous spring's TerraNova (CTB/McGraw-Hill, 2003) Total Math score was .51. *Double-Digit Subtraction* provides students with 3 min to complete 20 two-digit by two-digit subtraction problems with and without regrouping. The score is the number of correct answers. Percentage of agreement, calculated on 20% of protocols by two independent scorers, was 99.7. In the study presented here, coefficient alpha was .93. On a larger representative third-grade sample (Fuchs et al., 2003), criterion validity with the previous spring's TerraNova (CTB/McGraw-Hill) Total Math score was .48. In addition, we employed *Curriculum-Based Measurement Computation* (Fuchs, Hamlett, & Fuchs, 1990), a mixed-operations (addition and subtraction) one-page test displaying 25 items that sample the typical second-grade computation curriculum. Students have 3 min to complete as many problems as possible. The score is the number of problems correct. Staff entered responses into a computerized scoring program on an item-by-item basis, with 15% of tests re-entered by an independent scorer. Data-entry agreement was 99.7. Coefficient alpha on this sample was .93.

The Grade 3 Math Battery (Fuchs et al., 2003) includes one subtest of estimation. *Computational Estimation* was a one-page test displaying 20 two-digit addition problems, systematically representing problems for which answers were the sum of the tens, one 10-unit higher, or two 10-units higher. The tester reads a set of directions that include an explicated example: "Look at this problem ($23 + 45$). The estimated answer to this problem is 70. Write 70 under the problem. To estimate an answer, find the closest 10 to the exact answer. If I did the work to add this problem, the answer would be 68. But I don't want you to do the actual adding. Instead, I want you to estimate the answer to the nearest 10." Students had 5 min to do as many problems as possible. The score is the number of correct answers. Percentage of agreement, calculated on 20% of protocols by two independent scorers, was 99.8. In our study, coefficient alpha was .89. On a larger representative third-grade sample (Seethaler & Fuchs, 2006), criterion validity with the previous spring's TerraNova (CTB/McGraw-Hill, 2003) Total Math score was .58.

Other, Transfer Measures—The Grade 3 Math Battery (Fuchs et al., 2003) includes a subtest assessing *place value concepts*. Place Value Concepts incorporates 30 items across three pages. Four types of items vary whole-number places (tens, hundreds, thousands) and decimal places (none, tenths, hundredths). The first type of item presents a picture of blocks; the student writes the numeral or fills in the number of units for each place value. The second type of item presents a quantity for each place value (e.g., 6 tens 3 ones); the student writes the numeral. The third type of item presents a numeral; the student specifies the quantity for each place value. The fourth type of item presents a number; the student specifies which numeral is in which place value. The score is the number of correct answers. Percentage of agreement, calculated on 20% of protocols by two independent scorers, was 99.4. Coefficient alpha on this sample was .86. Although place value concepts for two-digit numbers, using pictorial representations of blocks, were addressed in procedural/estimation skill tutoring, only 6 (25%) of 24 items tapped this aligned content. Thus, we deemed this measure to demand transfer.

The Key Math-Revised (Connolly, 1998) Numeration subtest assesses mathematics concepts such as counting, correspondence, sequencing numbers, and ordinal positions with 24 items. Test items are ordered by difficulty; basal and ceiling rules are used. As reported in the

manual, alternate form reliability is .75, and the correlation with the Total Mathematics Score of the Iowa Test of Basic Skills (Hoover, Hieronymous, Dunbar, & Frisbie, 1993) is .67. Of the 24 items in Form A, 3 items across the entire grade span of the test (kindergarten through Grade 12) address place value concepts.

Following Jordan and Hanich (2000, adapted from Carpenter & Moser, 1984; Riley & Greeno, 1988; Riley, Greeno, & Heller, 1983), *Single-Digit Story Problems* comprises 14 story problems involving sums or minuends of 9 or less, reflecting change, combine, compare, and equalize relationships. The tester reads each item aloud; students have 30 sec to respond and can ask for rereading(s) as needed before the tester reads the next item. The score is the number of correct answers. A second scorer independently rescored 20% of protocols, with agreement of 99.8%. Coefficient alpha on this sample was .83. On a larger representative third-grade sample (Fuchs et al., 2006), criterion validity with the previous spring's TerraNova (CTB/McGraw-Hill, 2003) Total Math score was .66.

Double-Digit Story Problems—(Fuchs & Powell, 2003) comprises six story problems with double-digit operands (two problems do not require regrouping; four problems do require regrouping). The problems reflect change, combine, compare, and equalize relationships. The tester reads each item aloud; students complete their response before the tester reads the next item. Students can ask for rereading(s). The score is the number of correct answers. A second scorer independently rescored 20% of protocols, with agreement of 100%. Coefficient alpha on this sample was .73.

Tutoring

For each of the four conditions, tutoring involved 45 sessions, 3 sessions per week for 15 weeks. Tutors worked with individual students. The 22 tutors were (a) 4 full-time project coordinators (the 2 in Nashville were certified, experienced primary-grade teachers; the 2 in Houston, although not certified, had substantial tutoring experience) and (b) 18 part-time research assistants. In Nashville, research assistants were master's or doctoral students; in Houston, they were experienced teachers and tutors from the community. Each tutor was assigned a caseload that included all four tutoring conditions. Tutors were trained in 2 full days. Then, during the following week, the tutors studied the tutoring scripts (which were not read) and practiced implementing the procedures alone and with each other during the following week. Next, they conducted a session in each condition with a project coordinator who provided corrective feedback. Project coordinators also met with research assistants every 2 to 3 weeks to address problems or questions as they arose.

Fact Retrieval Skill Tutoring—Fact retrieval tutoring comprised three activities: computer-assisted instruction (7.5 min), flash card practice (4 min with corrective feedback), and cumulative review (4 min with corrective feedback), for a total of 15 to 18 min per session with time for transitions between activities.

Computer-assisted instruction, the first activity of each session, was conducted using Math Flash (Fuchs, Hamlett, & Powell, 2004), while the tutor supervised and answered questions. With Math Flash, the student was presented with an addition or subtraction math fact with the answer; the math fact “flashed” on the screen for 1.3 sec. When the math fact disappeared from the screen, the student used the computer keyboard to type the math fact from short-term memory. As the student typed the math fact, a number line illustrated the math fact at the top of the screen. This number line included 20 uncolored boxes, with a red line marking the perimeter of the first 10 boxes. As the student typed the first addend, boxes on the number line automatically turned blue to represent the quantity; as the student typed the second addend, boxes on the number line automatically shaded yellow to signify that

quantity. As the student typed a subtraction problem, boxes on the number line automatically shaded yellow to represent the minuend; black Xs were drawn through the yellow boxes to represent the subtrahend. After typing the math fact, the student pressed the return key to determine whether the math fact had been typed correctly. If correct, a numeral from 1 to 5 sparkled, and the student heard applause. After the first correct response, the numeral 1 sparkled; after the second correct response, the numeral 2 sparkled; and so on until five correct responses accumulated, at which point a picture of a “treasure,” such as a puppy or cake, dropped into the student’s “treasure box.” Then, the count from 1 to 5 began again, with a new “treasure” deposited into the treasure box after every 5 correct responses. The student tried to earn as many treasures as possible during the session (no concrete rewards were provided). If incorrect, the math fact reappeared on the screen, and remained while the student typed the math fact correctly (no numbers sparkled; no applause was heard). After feedback occurred, the student pressed the space bar to continue to the next math fact. At the end of 7.5 min, the program ended. Applause sounded while the student’s high score for the day and the score for that particular session were displayed. The student’s highest mastered score was recorded.

The second activity in each session was flash card practice, which included two types of flash cards. The first showed written math facts without answers. The student responded by saying the answers. The student had 2 min to respond to as many cards as possible, and then the student graphed the number correct each session. After mastery was demonstrated (i.e., three consecutive sessions with at least 35 correct responses), practice with the second set of flash cards began. With this set, the student was presented with a number line illustrating a math fact problem (as represented on Math Flash). The student stated the math fact represented by the number line. Again, the student had 2 min to respond to as many cards as possible, and then the student graphed the number correct each session. With both flash card procedures, the teacher corrected five incorrect responses with scripted remediation.

Cumulative review, the final activity in each session, was a paper–pencil activity. The student had 2 min to complete 15 math fact problems on paper. The teacher corrected the math problems out loud while the student observed. The daily score was recorded.

Procedural/Estimation Skill Tutoring—Procedural/estimation skill tutoring comprised three activities: computer-assisted instruction (5–10 min), flash card practice (4 min with corrective feedback), and cumulative review (4 min with corrective feedback), for a total of 15 to 18 min per session with time for transitions between activities. Intervention targeted procedural computation of two-digit numbers with and without regrouping and estimation of addition and subtraction of two-digit numbers with and without regrouping. Intervention addressed two-digit addition in Sessions 1 to 3, two-digit addition with and without regrouping in Sessions 4 to 9, two-digit subtraction without regrouping in Sessions 10 to 13m, two-digit subtraction with regrouping in Session 14, and two-digit addition and subtraction, with and without regrouping in Sessions 15 to 45.

Computer-assisted instruction, the first activity of each session, was conducted with Magic Math (Fuchs, Hamlett, & Powell, 2004), while the tutor supervised and answered questions. The Magic Math computer-assisted session comprised three segments, with one computer screen dedicated to each segment. Magic Math’s first segment addressed conceptual underpinnings, using pictorial representations of ones and tens. (In a series of initial sessions, manipulatives were used to concretize place value concepts.) The student was presented with an addition or subtraction problem. The student used an external computer mouse to represent ones and tens with pictures for the first number of the problem (addend or minuend). The pictorial representations of the ones and tens were at the top of the screen, and the student moved these into the ones and tens boxes to represent the first number. The

student then clicked on a checkmark next to the boxes to see if he or she had moved the correct quantity of ones and tens. If the work was correct, a character called the “Addition Magician” (for addition problems) or the “Subtraction Sorceress” (for subtraction) appeared and provided positive feedback via videotaped segments. If the work was incorrect, the Addition Magician or Subtraction Sorceress provided corrective feedback via videotaped segments to address the specific error the student had made and provided the student an opportunity to retry that step of the problem (i.e., the computer has been programmed to recognize different categories of errors and to provide support for the student to retry that step of the problem. The corrective feedback provided a hint and often included visual aides. This corrective feedback recurred for each step described next). Next, the student represented the ones and tens for the second number (if addition) or took away the ones and tens of the second number (if subtraction). The student clicked on the checkmark to indicate that he or she had completed this step, and the Addition Magician or Subtraction Sorceress appeared again and provided feedback. If the problem was addition and regrouping, the student, in the next step, moved 10 ones into a special box and clicked on the word *Glue*. The “Glue” button glued the 10 individual ones into 1 ten. If the problem was subtraction and regrouping, the student moved 1 ten into a special box and clicked on the word *Hammer*. The “Hammer” button broke the 1 ten into 10 ones. The student clicked on the checkmark to indicate that he or she had completed this step, and the Addition Magician or Subtraction Sorceress appeared again and provided feedback. In the next step, the student then moved the accumulated (for addition) or remaining (for subtraction) ones and tens to the corresponding boxes at the bottom of the screen. The student clicked on the checkmark to indicate that he or she had completed this step, and the Addition Magician or Subtraction Sorceress appeared again and provided feedback. Finally, the student counted the ones and tens in the bottom row of boxes and used the mouse to write the answer to the problem and clicked to indicate that the final step of the problem had been completed. Again, if the student made a mistake, the characters walked through the incorrect part of the problem (via videotaped segments), and the student got a chance to correct the work. When the student finished the first segment, the student clicked on the box at the bottom of the computer screen to move to the second segment.

Magic Math’s second segment taught procedural steps of two-digit addition and subtraction, relying on the same addition or subtraction problems worked in the first segment. The student was presented with nine questions/directives about the problem, which delineate a procedural routine for computing answers. The goal was for the student to use self-talk to internalize this task-analytic routine, represented by the nine questions/directives, for completing the procedural steps of the problem. After answering each question, the student clicked on a checkmark, and the Addition Magician or Subtraction Sorceress provided videotaped corrective feedback. The questions were as follows.

1. “What does the sign tell me to do?” The student clicked on “Add” or “Subtract.”
2. “Where do I start? Click on that column.” The student decided whether to start in the ones or tens column and clicked on that column.
3. “What plus what?” The student entered the numbers to be added or subtracted.
4. “Do I need to regroup?” The student clicked on “Yes” or “No.”
5. “Put my answer on the ones line.” The student calculated the answer to the ones column and wrote the answer.
6. “Where do I move? Click on that column.” The student clicked on the appropriate column.

7. “Now, what plus what?” The student entered the numbers to be added or subtracted in the tens column.
8. “Put my answer on the tens line.” The student wrote the answer to the tens column.
9. “What’s the answer?” The student wrote the complete answer to the problem.

After each question, the student clicked to receive videotaped corrective feedback, either praise or an explanation to match the student’s error.

Magic Math’s third segment focused on estimation. For estimation, the student worked on a double-digit addition problem that differed from the problems used in Segments 1 and 2. The addition problem was presented on the left side of the screen. The student was prompted to estimate the answer. If the student estimated correctly, he or she was finished with Magic Math’s computer-assisted activity for that day. If the student estimated incorrectly, the Addition Magician walked through a set of steps to teach estimation. The instructional sequence began by asking the student to “Add the tens.” The student added the two numbers in the tens column and marked an answer. The next step asked the student, “What are the three possible estimates?” The three correct possible estimates were the number the student had just calculated and the next two higher tens. The student clicked on those three numbers, and the first number turned blue. The second higher ten turned red, and the third higher ten turned green. The computer then said, “Add the ones. Mark it!” The student added together the two numbers in the ones column and marked the answer on a number line. The number line was color coded so that numbers 0 to 4 were blue, 5 to 14 were red, and 15 to 20 were green. The coloring of the number line helped the student determine whether the ones answer was closest to 0, 10, or 20 so the student would know whether to choose the lowest, middle, or highest estimate. The computer asked, “Is your estimate ... ?” and gave the student the three answer choices that he or she selected during the second step. After the student selected an answer, the computer prompted the student to write the estimated answer on the lines below the problem. The student clicked on the square at the bottom of the page to finish the day’s Magic Math session. After Session 15, the color coding disappeared. As with other segments, after each step just outlined, the student clicked on the checkmark to receive videotaped corrective feedback, either praise or an explanation to match the student’s error.

The second activity in each session was flash card practice, which included three types of flash cards. One set showed the student two-digit addition or subtraction problems with or without regrouping. The student responded by stating whether to add or subtract and then whether to regroup or not regroup. The student had 2 min to respond to as many cards as possible, and then the student graphed the number correct each session. After mastery was demonstrated (three consecutive sessions of 35 or more correct responses), practice with the second set of cards began, where the student was presented a card showing a two-digit addition problem, with or without regrouping. The student stated whether the sum of the ones column was closest to 0, 10, or 20. Again, the student had 2 min to respond to as many cards as possible, and then the student graphed the number correct each session. Once the student demonstrated mastery with this flash card activity, the student used the same set of cards (for the third flash card activity) to state the estimated answer to each two-digit addition problem. Again, the student had 2 min to respond, and then the student graphed the number of correct responses each session. For each flash card activity, the tutor corrected five incorrectly answered cards.

Cumulative review, the final activity in each session, was a paper–pencil activity. The student had 2 min to complete eight problems representing all problem types addressed in previous sessions. The tutor corrected the student’s work, talking through the procedure for

each correct and incorrect problem, while the student observed. The daily score was recorded.

Fact Retrieval + Procedural/Estimation Skill Tutoring—Fact retrieval + procedural/estimation skill intervention comprised four activities: Magic Math (5–10 min), Math Flash (7.5 min), the flash card practice incorporated in the procedural/estimation skill intervention (4 min with corrective feedback), and cumulative review (4 min with corrective feedback), for an approximate total of 25 min per session with time for transitions between activities. Note that to avoid expanding the time required for the combined fact retrieval + procedural/estimation skill intervention even further, we did not include the flash card practice employed with retrieval tutoring.

Word-Identification Skill Tutoring—Word-identification skill tutoring comprised two activities: computer-assisted instruction (7.5 min) and repeated reading (7 min with corrective feedback), for a total of 15 to 18 min per session with time for transitions between activities. Students were placed into the first-grade, second-grade, or third-grade material for word-identification intervention based on their grade-equivalent score on WRAT–Reading.

Computer-assisted instruction, the first activity of each session, was conducted with Reading Flash (Fuchs, Hamlett, Powell, & Seethaler, 2004), while the tutor supervised and answered questions. The student was presented with a word that flashed on the screen for 1.3 sec. When the word disappeared from the screen, the student used the computer keyboard to type in the word. The student pressed the Return key to see if he or she spelled the word correctly. The computer stated and spelled the word out loud for the student. If the word's spelling was correct, a number from 1 to 5 sparkled and the student heard applause. If the student's spelling was incorrect, the word reappeared and remained on the screen while the student typed the word correctly; no numbers sparkled and no applause was heard. The student pressed the space bar to continue to the next word. Once the student wrote five words correctly, a picture of a "treasure," such as a sun or monkey, dropped into the student's "treasure box" (no concrete rewards were provided). The student tried to earn as many treasures as possible during the session. At the end of 7.5 min, the program ended. Applause sounded while the student's highest mastered score (cumulative across sessions) and the score for that particular session were displayed. The student's highest mastered score was recorded.

The second activity in each session was repeated readings. The student was presented with a paper copy of a story from assigned reading level (determined by WRAT–Reading grade equivalent). The student had 2 min to read the story aloud. The teacher followed along on a computer that displayed the same story on the screen as well as a timer that flashed the amount of time the student had been reading. The teacher used the cursor to mark words read incorrectly by the student. If the student paused for more than 3 sec on any word, the teacher told the student the word, marked it as incorrect, and encouraged the student to keep reading. At the end of 2 min, the teacher marked the last word read. The student's correct number of words read during the 2 min appeared on the computer screen. The teacher shared the score with the student, and the student graphed the score. The student was then prompted to reread the story as the teacher again marked mistakes. As with the first reading, the teacher shared the score with the student, and the student graphed this score on the same graph. Then the student read the story for a third time as the teacher again marked mistakes. The student graphed the final score, and the teacher discussed the day's performance with comparisons to previous days' scores.

Treatment Fidelity

Every session in each tutoring condition was audiotaped. Four research assistants independently listened to tapes while completing a checklist to identify the percentage of essential points in that lesson. We sampled 16% of the tapes such that conditions, research assistants, and lesson types were sampled comparably. At the site where the tutoring protocols had been developed, Nashville, the mean percentage of points addressed was 99.72 ($SD = 0.32$) for the fact retrieval skill intervention, 99.69 ($SD = 0.44$) for the procedural/estimation skill intervention, 99.63 ($SD = 0.43$) for the fact retrieval + procedural/estimation skill intervention, and 99.71 ($SD = 0.45$) for the word-identification condition. In Houston, the mean percentage of points addressed was 99.82 ($SD = 0.24$) for the fact retrieval skill intervention, 99.69 ($SD = 0.53$) for the procedural/estimation skill intervention, 99.39 ($SD = 0.68$) for the fact retrieval + procedural/estimation skill intervention, and 99.79 ($SD = 0.24$) for the word-identification condition. Tutors also recorded the duration of each session. In Nashville, the total amount of intervention time in minutes averaged 632.39 ($SD = 41.97$) for the fact retrieval skill intervention, 681.28 ($SD = 86.87$) for the procedural/estimation skill intervention, 984.22 ($SD = 106.00$) for the fact retrieval + procedural/estimation skill intervention, and 728.00 ($SD = 55.28$) for the word-identification intervention. In Houston, the total amount of intervention time in minutes averaged 719.64 ($SD = 40.68$) for the fact retrieval skill intervention, 816.82 ($SD = 154.49$) for the procedural/estimation skill intervention, 1150.50 ($SD = 136.64$) for the fact retrieval + procedural/estimation skill intervention, and 791.50 ($SD = 62.48$) for the word-identification intervention. An analysis of variance revealed a significant effect, as expected, for condition, $F(3, 120) = 107.72, p < .001$. Follow-ups indicated the following: combined > the other three conditions, procedural/estimation skill intervention > fact retrieval skill intervention, word identification skill intervention > fact retrieval skill intervention. In addition, there was a significant effect for site, $F(1, 120) = 40.91, p < .01$. Houston tutors took more time than did Nashville tutors; however, the interaction between condition and site was not significant, $F(1, 120) = 1.48, p = .223$.

Procedure

Screening to identify students who met MD and MDRD criteria occurred in September using WRAT–Arithmetic (in large groups), WRAT–Reading (individually), and WASI (individually). During October, pretest data were collected: Addition Fact Fluency 0–8, Subtraction Fact Fluency 0–8, Addition Fact Fluency 0–12, Subtraction Fact Fluency 0–12, Addition Fact Fluency 0–18, Subtraction Fact Fluency 0–18, Double-Digit Addition, Double-Digit Subtraction, Curriculum-Based Measurement Computation, Computational Estimation, Place Value Concepts, KeyMath Numeration, Single-Digit Story Problems, and Double-Digit Story Problems. These measures were administered in small groups (except when scheduling dictated one-to-one administration or for makeups and except KeyMath Numeration, which was administered individually as dictated by standard procedures). For Single-Digit Story Problems and Double-Digit Story Problems, each story problem was read aloud, with research assistants providing time for students to complete work before progressing to the next item. Intervention began the 1st week of November and ran through the 2nd week of March. Posttesting occurred during the 3rd and 4th week in March, using the same measures as had been used at pretest, except that we also administered Double-Digit Story Problems. Trained research assistants collected data using standardized directions.

RESULTS

Analyses

Preliminary analyses of the 132 students (at pretest) and 128 students (at posttest) included distributional exploration of relevant variables via statistical (e.g., skewness, kurtosis) and graphical (e.g., box plots, stem and leaf plots) means. One student with MDRD in the fact retrieval + procedural/estimation skill tutoring group was found to be a low outlier on multiple measures at pre- and posttest and was excluded from further analyses. No other consistently high or low outliers were identified; therefore, analyses were completed with 127 students. Most measures were normally distributed. However, the Addition Fact Fluency 0–8 tests suffered a ceiling effect, even at pretest; so we eliminated it. Computational Estimation produced a U-shaped distribution, but because no transformation resulted in significant improvement, the original scores were retained. To decrease Type I error inflation and to enhance construct coverage, measures were grouped according to type of math skill, and a factor score with a mean of 0 ($SD = 1$) was derived for each of type of math skill: Fact Retrieval Skill (five measures: Subtraction Fact Fluency 0–8, Addition Fact Fluency 0–12, Subtraction Fact Fluency 0–12, Addition Fact Fluency 0–18, Subtraction Fact Fluency 0–18), Procedural Computational Skill (three measures: Double-Digit Addition, Double-Digit Subtraction, and Curriculum-Based Measurement Computation), Story Problems (one measure at pretest: Single-Digit Story Problems; two measures at posttest: Single-Digit Story Problems and Double-Digit Story Problems), and Math Concepts (two measures: Place Value Concepts and KeyMath Numeration). Computational Estimation Skill addressed a fifth construct but was identified by a single outcome (Computational Estimation), and so a z -standardized transformation was used.

The factors of interest were disability status (MD vs. MDRD) and tutoring condition (fact retrieval tutoring vs. procedural/estimation tutoring vs. combined [fact + procedural/estimation skill tutoring] vs. word identification tutoring). Therefore, the primary analyses utilized a two-way analysis of covariance, with the four tutoring levels and two disability status levels comprising the factors and with pretest performance used as the covariates. The first step involved testing the relation of the pretest covariates to the factors at posttest, which was not significant for any outcome; therefore, these were trimmed from the models. Next, the interaction of tutoring condition and disability status was tested and, if not significant, was trimmed from the final model. All omnibus models were significant, although we do not provide these results to highlight the significant unique contributions. In the case of significant main effects or interactions, follow-up comparisons were conducted on the adjusted posttest means using a Tukey correction for multiple comparisons. Table 3 displays posttest data for the outcome factors by tutoring condition and by difficulty status.

Finally, effect sizes were computed for all comparisons of interest. The most basic formula for computerized effect size of the differences between groups is given by Cohen's $d = ([\text{Mean } 1 - \text{Mean } 2]/SD)$. Effect sizes can be computed in different ways, but all essentially represent the degree of difference between groups in the context of the variability within groups. Methods of computing effect size vary in terms of which numerator differences are chosen (e.g., unadjusted posttest means vs. adjusted least-squares means) or in terms of which measure of variability is utilized (e.g., the standard deviation of the control group, the pooled standard deviation across groups, the root mean square error term, etc.). We opted for unadjusted group means in the numerator and the pooled standard deviation across the groups being compared in the denominator to correct for sample overestimation bias (Hedges & Olkin, 1985), which was generally small. Effect sizes were similar when adjusted least-squares means were used in place of the unadjusted factor score means. Table 4 displays effect size values for marginal means for tutoring condition and for difficulty status.

Fact Retrieval Skill

There were no interactions of pretest with tutoring condition or difficulty status, and the interaction of tutoring condition with difficulty status was not significant, $F(3, 118) = < 1, p > .05$. In the final model, there was an effect of pretest, $F(1, 121) = 42.63, p < .0001$, and there was a significant main effect for tutoring condition, $F(3, 121) = 5.35, p < .002$, but not for difficulty status, $F(1, 121) = 2.23, p > .05$. Examination of adjusted means and follow-up analyses with Tukey correction for multiple comparisons revealed that fact retrieval tutoring outperformed each of the other conditions, including word identification tutoring ($p < .003, d = 0.78$), procedural/estimation tutoring ($p < .02, d = 0.69$), and combined tutoring ($p < .01, d = 0.89$). Adjusted means were not significantly different for students with MD versus MDRD, as noted, although the unadjusted effect size was 0.43, favoring students with MD. Thus, as expected, the group that was tutored on fact retrieval outperformed the other groups on the aligned outcome measure.

Procedural Computation Skill

There were no interactions of pretest with tutoring condition or difficulty status, and the interaction of tutoring condition with difficulty status was not significant, $F(3, 118) = < 1, p > .05$. In the final model, there was an effect of pretest, $F(1, 121) = 69.68, p < .0001$. The main effect for treatment was not significant, $F(3, 121) < 1, p > .05$. Thus, in contrast to expectations, the groups that received training in computational procedures did not outperform the other groups. The main effect for difficulty status approached significance, $F(1, 121) = 3.21, p < .08$ (the unadjusted effect size was 0.48, favoring students with MD over MDRD).

Computational Estimation Skill

There were no interactions of pretest with tutoring condition or difficulty status, and the interaction of tutoring condition with difficulty status was not significant, $F(3, 118) = < 1, p > .05$. In the final model, there was an effect of pretest, $F(1, 121) = 8.24, p > .005$, and there was a significant main effect of tutoring condition, $F(3, 121) = 11.19, p < .0001$, but not of difficulty status, $F(1, 121) = 2.18, p > .05$. Examination of adjusted means and follow-up analyses with Tukey correction for multiple comparisons revealed that procedural/estimation skill tutoring outperformed two other groups: word identification tutoring ($p < .0001, d = 1.47$) and fact retrieval tutoring ($p < .0005, d = 1.03$); in addition, students in combined tutoring also outperformed those in word identification tutoring ($p < .004, d = 0.88$). No other differences were noted. Altogether, the groups that received training in estimation outperformed the fact retrieval and word identification groups, but the effect was less apparent for the combined group.

Story Problems

There were no interaction of pretest with tutoring condition or difficulty status, and the interaction of tutoring condition with difficulty status was not significant, $F(3, 117) = < 1, p > .05$. In the final model, there was an effect of pretest, $F(1, 120) = 86.68, p < .0001$, but main effects for tutoring condition and difficulty status were not significant, $F(3, 120) < 1, p > .05$, and $F(1, 120) < 1, p > .05$, respectively. For students with MD versus MDRD, the unadjusted effect size was 0.61 favoring students with MD. Thus, no generalization is apparent on unaligned mathematics measures.

Math Concepts

There were no interactions of pretest with tutoring condition or difficulty status, and the interaction of tutoring condition with difficulty status was not significant, $F(3, 118) = < 1, p > .05$. In the final model, there was an effect of pretest, $F(1, 121) = 118.99, p < .0001$, but

main effects for tutoring condition and difficulty status were not significant, $F(3,121) < 1$, $p > .05$, and $F(1, 121) = 1.68$, $p > .05$, respectively. For students with MD versus MDRD, the unadjusted effect size was 0.69 favoring students with MD. Thus, no generalization is apparent on unaligned mathematics measures.

DISCUSSION

The primary purposes of this study were to assess the efficacy of tutoring to remediate third-grade students' mathematics deficits and to explore whether remediation is differentially efficacious depending on whether students experienced mathematics difficulty alone or in combination with reading problems. In addition, we examined the transferability of the tutoring protocols by assessing whether tutoring efficacy interacted with site (i.e., whether intervention occurred at a site local to the intervention developers as opposed to a distal site). In terms of this third purpose, analyses revealed no significant interactions involving site that affected the interpretation of results, leading us to conclude that the tutoring protocols were transportable (and to trim site from the analyses). The transportability of the tutoring protocols suggests the potential for scaling up these tutoring protocols, given the proviso that tutors are trained as done in our study. That is, tutors (a) receive 2 days of training, (b) practice implementing the procedures alone and with each other during the subsequent week, (c) conduct a session in each condition with a supervisor who provides corrective feedback, (d) study (not read) the tutoring scripts as they implement tutoring, and (e) meet with fellow tutors and the supervisor every 2 to 3 weeks to address problems or questions as they arise.

With respect to the first purpose (efficacy of the tutoring procedures), findings generally support the notion that tutoring effects are specific to the skill where intervention occurs, with limited transfer to related aspects of mathematical cognition—at least for students who have accrued substantial mathematics deficits by the beginning of third grade. On the fact retrieval skill outcome, students in the fact retrieval skill tutoring condition outperformed students in the procedural computation/computational estimation tutoring condition and outperformed students in the word identification tutoring condition. Respective effect sizes were in the moderate to large range: 0.69 to 0.78. Improvement in fact retrieval skill in this population is notable because fact retrieval deficits are considered a signature deficit of students with math disability and are thought to be a precursor to and partial explanation for difficulties with other math skills (e.g., Fuchs, Fuchs, Compton, et al., 2006). The specificity of effects is also notable. That is, sustained, intensive tutoring on procedural computation/computational estimation did not result in significantly enhanced fact retrieval skill, as demonstrated by the fact that the effect size relative to word identification tutoring on the fact retrieval outcome was a modest 0.17.

It is also interesting to consider that the tutoring condition that combined fact retrieval with procedural computation/computational estimation did not enhance fact retrieval performance, with a disappointing effect size of -0.04 . This is probably because the combined tutoring condition sacrificed instructional time on fact retrieval (specifically, the tutor-directed portion of fact retrieval practice) as a strategy for minimizing differences in instructional time between the single-versus combined-focus tutoring conditions. This suggests the need for sufficient teacher-directed practice on math facts to produce meaningful improvements in fact retrieval performance among this population of learners. Results extend the existing literature (e.g., Christensen & Gerber, 1990; Okolo, 1992; Tournaki, 2003) by demonstrating specific effects on fact retrieval within the context of a randomized controlled trial that (a) was conducted across local and distal sites, (b) focused on students with documented low performance in math or in math and reading, (c) included

a control for instructional time and attention, and (d) incorporated research-based instructional design principles across study conditions.

In a similar way, effects on computational estimation were convincing and specific. For this aspect of mathematical cognition, effects were statistically significant and large only for conditions that specifically focused on procedural computation/computational estimation. This is interesting because skill with number combinations, as reflected in fact retrieval tutoring, is clearly requisite to computational estimation. Also, effects favoring the procedural computation/computational estimation condition occurred regardless of whether tutoring incorporated a single focus on procedural computation/computational estimation skill or a combined focus across procedural computation/computational estimation and fact retrieval. Effect sizes for both conditions were large compared to the word recognition control condition, with effect sizes for the single and combined focus of 1.47 and 0.88. Effect sizes were also respectively large when comparing the single-focus procedural computation/computational estimation skill tutoring condition against the fact retrieval tutoring condition (effect size = 1.03).

There was some suggestion that combined tutoring resulted in a diminished effect. For example, when comparing the combined condition against the single-focus procedural computation/computational estimation condition, the effect size on the computational estimation outcome was -0.55 . This is surprising given that the procedural computation/computational estimation tutoring component was identical across the single-focus and combined conditions and that the combined tutoring sessions were in fact longer than the single-focus sessions (an average total minutes of instruction of 1,054 vs. 713). Although the computational skill outcomes between these two conditions were not statistically significant, the effect size of -0.55 suggests that the attentional difficulties of students with mathematics difficulties may render longer tutoring sessions less efficacious (Fuchs et al., 2005; Fuchs, Fuchs, Compton, et al., 2006). This warrants future study with larger samples. Although it is tempting to attribute the effects on computational estimation outcomes to the computational estimation component of tutoring, we note that our study provides no basis for determining whether effects on computational estimation accrued because of the procedural computation component or the computational estimation component. Future work should address this issue. At the same time, few studies have examined the efficacy of methods for remediating deficits in computational estimation, and our findings do validate an approach for accomplishing this important goal.

In terms of procedural computation performance, results were disappointing. Procedural computation/computational estimation tutoring, with or without fact retrieval tutoring, failed to render significant effects on the procedural computation outcome. It is difficult to explain this lack of effect, when the procedural computation component of the tutoring addressed the conceptual basis for addition and subtraction (with and without regrouping) and incorporated a step-by-step task analysis of the procedure for completing two-digit addition and subtraction problems (with and without regrouping). It is possible that the computer medium (where the bulk of instruction occurred) required transfer to paper, which did not occur sufficiently. This suggests the need for additional study comparing computer-mediated instruction against other delivery options for this population of learners. Even so, effect sizes for the fact retrieval and procedural computation conditions were, respectively, .44 and .30, suggesting the possibility that significant effects would emerge for these existing tutoring protocols if sample size were increased.

Difficulty in transferring knowledge and skills is a well-established characteristic of students with disabilities (e.g., Cooper & Sweller, 1987; Fuchs, Fuchs, Karns, Hamlett, & Katzaroff, 1999; Mayer, 1998; Woodward & Baxter, 1997), and the specificity of effects in our

findings represents further evidence of this difficulty. That is, when remediation focused (with sufficient tutor-directed practice opportunities) on automatic retrieval of number combinations, we observed differential outcomes on fact retrieval but not on procedural computation or computational estimation. When remediation focused on computational estimation, stronger outcomes were evident on computational estimation but not on fact retrieval. Although these findings provide a strong demonstration of treatment specificity and are therefore highly interpretable, it is disappointing that students did not transfer across these two related areas of mathematical cognition.

We saw additional evidence of failure to transfer in the story problem and math concepts outcomes, where statistically significant tutoring condition effects did not occur. More optimistically, however, we do note that on story problems, the fact retrieval tutoring condition performed 0.37 standard deviations better than the control word identification tutoring condition, and the procedural computation/computational estimation condition scored 0.31 standard deviations ahead of the control word identification condition. Moreover, on math concepts, students in the procedural computation/computational estimation condition (which addressed place value concepts, one type of item included on the math concepts measures) scored 0.35 standard deviations ahead of the control group and 0.31 standard deviations ahead of the fact retrieval condition. These findings, which tentatively suggest that modest transfer may have occurred, should be pursued in future work with larger samples.

The second major purpose of our study was to explore whether remediation is differentially efficacious depending on whether students experience mathematics difficulty alone versus in combination with reading problems. As shown in Table 2, we observed clear pretest advantages for MD students over MDRD students on every measure except computational estimation. Yet results revealed comparable responsiveness to intervention. The absence of significant interactions between students' difficulty status and tutoring condition does not lend support to the proposition that MD and MDRD constitute distinct subtypes of mathematics disability. Of course, demonstrating the null effect does not provide the basis for strong conclusions. The lack of significant interactions between tutoring condition and difficulty status may have occurred for a host of reasons. For example, differential response may be revealed with different interventions or with more stringent definitions of difficulty status. Consequently, additional research should continue to explore the issue of differential responsiveness with randomized controlled trials.

Acknowledgments

This research was supported in part by Grant No. 1 P01046261 and Core Grant No. HD15052 from the National Institute of Child Health and Human Development to Vanderbilt University. Statements do not reflect the position or policy of these agencies, and no official endorsement by them should be inferred.

REFERENCES

- Aiken LR. Language factors in learning mathematics. *Review of Educational Research*. 1972; 42:359–385.
- Badian, NA. Dyscalculia and nonverbal disorders of learning. In: Myklebust, HR., editor. *Progress. Grune & Stratton*; New York: 1983. p. 235-264.
- Carpenter TP, Moser JM. The acquisition of addition and subtraction concepts in grades one through three. *Journal of Research in Mathematics Education*. 1984; 15:179–203.
- 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 MA, Fuchs L, Fletcher JM. Cognitive arithmetic differences in learning disability groups and the role of behavioral inattention. *Learning Disabilities Research & Practice*. 2007; 22:25–35.
- Clements DH, Sarama J. Effects of preschool mathematics curriculum: Summary of research on the Building Blocks project. *Journal for Research in Mathematics Education*. 2007; 38:136.
- Compton DL, Fuchs LS, Fuchs D. The course of reading and mathematics disability in first grade: Identifying latent class trajectories and early predictors. 2006 Manuscript submitted for publication.
- Connolly, AJ. KeyMath–Revised. American Guidance Service; Circle Pines, MN: 1998.
- Cooper G, Sweller J. Effects of schema acquisition and rule automation on mathematical problem solving transfer. *Journal of Educational Psychology*. 1987; 79:347–362.
- CTB. McGraw-Hill. TerraNova technical manual. Author; Monterey, CA: 2003.
- Delazer M, Bartha L. Transcoding and calculation in asphasia. *Asphasiology*. 2001; 15:649–679.
- Fletcher, JM.; Lyon, GR.; Fuchs, LS.; Barnes, MA. Learning disabilities: From identification to intervention. Guilford; New York: 2007.
- Fuchs LS, Fuchs D. Mathematical problem solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities*. 2002; 35:563–574. [PubMed: 15493253]
- Fuchs LS, 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, et al. The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*. 2006; 98:29–43.
- Fuchs LS, Fuchs D, Hamlett CL, Stecker PM. Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal*. 1991; 28:617–641.
- Fuchs LS, Fuchs D, Karns K, Hamlett CL, Katzaroff M. Mathematics performance assessment in the classroom: Effects on teacher planning and student learning. *American Educational Research Journal*. 1999; 36:609–646.
- Fuchs LS, Fuchs D, Prentice K. Responsiveness to mathematical problem-solving instruction among students with risk for mathematics disability with and without risk for reading disability. *Journal of Learning Disabilities*. 2004; 4:293–306. [PubMed: 15493402]
- Fuchs LS, Fuchs D, Stuebing K, Fletcher JM, Hamlett CL, Lambert WE. Problem solving and calculation skill: Shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology*. in press.
- Fuchs LS, Fuchs D, Yazdian L, Powell SR. Enhancing first-grade children's mathematical development with peer-assisted learning strategies. *School Psychology Review*. 2002; 31:569–584.
- Fuchs LS, Hamlett CL, Fuchs D. Curriculum-based measurement: Computation. 1990 Information available from L. S. Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37203.
- Fuchs LS, Hamlett CL, Powell SR. Grade 3 Math Battery. 2003 Available from L.S. Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37203.
- Fuchs LS, Hamlett CL, Powell SR. Math Flash and Magic Math. 2004 Unpublished computer programs; information available from L.S. Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37203.
- Fuchs LS, Hamlett CL, Powell SR, Seethaler PM. Reading Flash. 2004 Unpublished computer program; information available from L. S. Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37203.
- Fuchs LS, Powell SR. Double-Digit Story Problems. 2003 Information available from L. S. Fuchs, 328 Peabody, Vanderbilt University, Nashville, TN 37203.
- Griffin, SA.; Case, R.; Siegler, RS. Rightstart: Providing the central conceptual prerequisite for first formal learning of arithmetic to students at risk for school failure. In: McGilly, K., editor. *Classroom lessons: Integrating cognitive theory and classroom practice*. MIT Press; Cambridge, MA: 1994. p. 25-50.

- Hanich LB, Jordan NC, Kaplan D, Dick J. Performance across different areas of mathematical cognition in children with learning disabilities. *Journal of Educational Psychology*. 2001; 93:615–626.
- Hedges, LV.; Olkin, I. Statistical methods for meta-analysis. Academic; Orlando, FL: 1985.
- Hoover, HD.; Hieronymous, AN.; Dunbar, SB.; Frisbie, DA. Iowa Test of Basic Skills, Form K. Riverside; Itasca, EL: 1993.
- Jordan NC, Hanich L. Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities*. 2000; 33:567–578. [PubMed: 15495398]
- Jordan NC, Hanich LB, Kaplan D. Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology*. 2003; 85:103–119. [PubMed: 12799164]
- Landerl K, Began A, Butterworth B. Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students. *Cognition*. 2004; 93:99–125. [PubMed: 15147931]
- Lewis C, Hitch GJ, Walker P. The prevalence of specific arithmetic difficulties and specific reading difficulties in 9- to 10-year-old boys and girls. *Journal of Child Psychology and Psychiatry*. 1994; 35:283–292. [PubMed: 8188799]
- Mayer DP. Do new teaching standards undermine performance on old tests? *Educational Evaluation and Policy Analysis*. 1998; 15:1–16.
- 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.
- Psychological Corporation. Wechsler Abbreviated Scale of Intelligence. Harcourt Brace; San Antonio, TX: 1999.
- Rasanen P, Ahonen T. Arithmetic disabilities with and without reading difficulties: A comparison of arithmetic errors. *Developmental Neuropsychology*. 1995; 11:274–295.
- Riley MS, Greeno JG. Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction*. 1988; 5:49–101.
- Riley, MS.; Greeno, JG.; Heller, JI. The development of mathematical thinking. Academic; New York: 1983. Development of children's problem-solving ability in arithmetic; p. 153-196.
- 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.
- Seethaler PM, Fuchs LS. The cognitive correlates of computational estimation skill among third-grade students. *Learning Disabilities Research & Practice*. 2006; 21:233–243.
- Shalev RS, Auerbach J, Manor O, Gross-Tsur V. Developmental dyscalculia: Prevalence and prognosis. *European Child & Adolescent Psychology*. 2000; 9(6):58–65.
- Shaywitz SE, Fletcher JM, Holahan JM, Shneider AE, Marchione KE, Stuebing KK, et al. Persistence of dyslexia: The Connecticut Longitudinal Study at adolescence. *Pediatrics*. 1999; 104:1351–1359. [PubMed: 10585988]
- 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]
- Wilkinson, GS. Wide Range Achievement Test 3. Wide Range; Wilmington, DE: 1993.
- Woodward J, Baxter J. The effects of an innovative approach to mathematics on academically low-achieving students in inclusive settings. *Exceptional Children*. 1997; 63:373–388.

Table 1

Demographic and screening data by tutoring condition and by difficulty status

Variable	Tutoring condition				Difficulty status	
	Fact retrieval ^a	Procedural comp/estimation ^b	Combined ^c	Word identification ^d	MD ^e	MDRD ^f
Age in years	9.48 (0.7)	9.36 (0.4)	9.28 (0.5)	9.25 (0.6)	9.16 (0.4) <i>a</i>	9.52 (0.6) <i>b</i>
Female	44%	28%	52%	60%	45%	48%
Subsidized lunch	66%	69%	84%	66%	59%	83%
Special education	16%	24%	16%	17%	5%	32%
Retained	22%	21%	19%	17%	9%	30%
English as a second language	9%	10%	10%	14%	11%	11%
Ethnicity AA	38%	41%	58%	60%	41%	59%
White	28%	31%	19%	17%	28%	19%
Hispanic	25%	28%	31%	17%	27%	17%
Other	9%	0%	3%	6%	5%	5%
WASI FSIQ	90.16 (11.3)	90.10 (12.4)	90.29 (10.7)	91.37 (10.7)	93.34 (11.4) <i>a</i>	86.62 (9.4) <i>b</i>
WRAT Arithmetic	83.94 (7.0)	83.48 (5.6)	82.97 (7.3)	85.51 (4.4)	86.13 (4.8) <i>a</i>	81.90 (6.7) <i>a</i>
WRAT Reading	93.00 (15.5)	95.03 (12.5)	93.10 (13.8)	92.69 (14.3)	105.03 (6.8) <i>a</i>	81.59 (8.3) <i>a</i>

Note. Percentages are computed relative to the number of individuals in that group (e.g., 44% of the 32 students in Fact Retrieval Tutoring were female). Percentages for Ethnicity within a column total of 100%. Numbers in parentheses are standard deviations. For mean scores, values in a given row with different subscripts are significantly different from one another. Although the difference between math difficulty alone (MD) and math and reading difficulty (MDRD) appears different for Wide Range Achievement Test-III (WRAT) Reading, the analyses controlled for reading ability, rendering the effect nonsignificant. AA = African American; WASI FSIQ = Wechsler Abbreviated Scales of Intelligence Full Scale IQ.

^a *n* = 32.

^b *n* = 29.

^c *n* = 31.

^d *n* = 35.

^e *n* = 64.

^f *n* = 63.

Table 2

Pretest performance by tutoring condition and difficulty status

Variable	Tutoring condition					Difficulty status		
	F	Fact retrieval ^a	Procedural Comp/estimation ^b	Combined ^c	Word-identification ^d	F	MD ^e	MDRD ^f
Fact retrieval	<1	0.16 (1.02)	0.10 (1.05)	-0.23 (0.84)	-0.03 (1.07)	6.18 [*]	0.21 (1.00)	-0.22 (0.96)
Procedural comp	1.28	0.16 (1.12)	0.15 (0.86)	-0.27 (1.02)	-0.03 (1.07)	4.59 [*]	0.18 (0.99)	-0.19 (0.98)
Comp estimation	1.38	0.16 (1.01)	0.18 (1.22)	-0.05 (0.95)	-0.25 (0.80)	<1	-0.01 (0.93)	0.01 (1.08)
Story problems	2.76 [*]	0.13 (0.99)	0.29 (0.91)	-0.30 (1.02)	-0.09 (1.02)	25.71 ^{***}	0.41 (0.82)	-0.41 (1.00)
Math concepts	<1	0.08 (0.90)	0.18 (0.82)	-0.21 (1.16)	-0.04 (1.07)	17.05 ^{**}	0.34 (0.71)	-0.35 (1.13)

Note. Most scores are factor scores derived from a combination of similar measures. See text for explanation. Story Problems and Computational Estimation, however, are individual pretest variables, which have been z-score standardized for comparative purposes. MD = math difficulty alone; MDRD = math and reading difficulty; comp = computation.

Other F values not significant. For story problems the significant main effects are superseded by a significant interaction of tutoring condition and difficulty status at pretest, $F(3, 118) = 3.62, p < .05$, partial $\eta^2 = 0.09$.

^a $n = 32$.

^b $n = 27$.

^c $n = 31$.

^d $n = 35$.

^e $n = 64$.

^f $n = 63$.

* $p < .05$.

** $p < .0001$.

Table 3

Posttest performance by tutoring condition and difficulty status

Variable	<i>F</i>	Tutoring condition			Difficulty status		
		Fact retrieval ^a	Procedural comp/estimation ^b	Combined ^c	Word identification ^d	MD ^e	MDRD ^f
Fact retrieval	5.35*	0.57 (0.94)	-0.16 (0.86)	-0.27 (0.92)	-0.23 (1.07)	2.23	0.21 (0.97)
Procedural comp	<1	0.25 (0.93)	0.14 (0.99)	-0.19 (1.04)	-0.17 (1.01)	3.21	0.24 (0.92)
Comp estimation	11.19**	-0.23 (0.99)	0.69 (0.74)	0.24 (0.87)	-0.57 (0.93)	2.18	0.12 (0.99)
Story problems	<1	0.23 (1.07)	0.14 (0.90)	-0.18 (0.95)	-0.17 (1.04)	<1	0.29 (0.79)
Math concepts	<1	-0.03 (1.11)	0.23 (0.84)	-0.06 (0.99)	-0.11 (1.04)	1.68	0.33 (0.81)

Note. Most scores are factor scores derived from a combination of similar measures. See text for explanation. Computational Estimation, however, is a single variable, which has been z-score standardized for comparative purposes MD = math difficulty alone; MDRD = math and reading difficulty ; comp = computation.

Other *F* values not significant.

^a
n = 32.

^b
n = 27.

^c
n = 31.

^d
n = 35.

^e
n = 64.

^f
n = 63.

*
p < .05.

**
p < .0001.

Effect sizes by tutoring condition and difficulty status

Note. Effect sizes are based on unadjusted means at posttest and pooled standard deviation of both groups being compared. Effect sizes for diagnosis (MD v. MDRD) appear larger than suggested in Table 2, because significance in Table 2 is based on adjustment for treatment group. Effect sized indicate higher means for the first group named in the comparison, except those in bold, which indicate higher means for the second group names in the comparison. Values in parentheses are confidence intervals for the effect sizes. Proc = procedural; est = estimation; MD = math difficulty alone; MDRD = math and reading difficulty; comp = computation.

J Res Educ Eff. Author manuscript; available in PMC 2011 June 23.