The Effects of Cognitive Strategy Instruction on Math Problem Solving of Middle-School Students of Varying Ability

Article *in* Journal of Educational Psychology · May 2014

DOI: 10.1037/a0035176

4 authors, including:

Jennifer Krawec
University of Miami

Samantha Dietz
University of Miami

Journal of Educational Psychology
2014, Vol. 106, No. 2, 469–481

0022-0663/14/$12.00 DOI: 10.1037/a0035176

The Effects of Cognitive Strategy Instruction on Math Problem Solving of Middle-School Students of Varying Ability

Marjorie Montague and Jennifer Krawec
University of Miami

Craig Enders
Arizona State University

Samantha Dietz
University of Miami

The effects of a mathematical problem-solving intervention on students' problem-solving performance and math achievement were measured in a randomized control trial with 1,059 7th-grade students. The intervention, *Solve It!*, is a research-based cognitive strategy instructional intervention that was shown to improve the problem-solving performance of 8th-grade students with and without learning disabilities (LD). The purpose of the present study was to determine whether the effectiveness of the intervention could be replicated with younger students. Forty middle schools in a large urban school district were included in the study, with one 7th-grade math teacher participating at each school (after attrition, $n = 34$). *Solve It!* was implemented by the teachers in their inclusive math classrooms. Problem-solving performance was assessed using curriculum-based math problem-solving measures, which were administered as a pretest and then monthly over the course of the 8-month intervention. Students who received the intervention ($n = 644$) embedded in the district curriculum showed a significantly greater rate of growth on the curriculum-based measures than students in the comparison group ($n = 415$) who received the district curriculum only. Results of the Bayesian analyses indicated that the intervention effect was somewhat stronger for low-achieving students than for average-achieving students. Overall, findings from the present study as well as the previous study with 8th-grade students indicate that the intervention was effective across ability groups and is an appropriate program to use in inclusive classrooms with students of varying math ability.

Keywords: mathematics, problem solving, strategy instruction

Supplemental materials: http://dx.doi.org/10.1037/a0035176.supp

Students of all abilities, but particularly students with learning disabilities (LD), identify math problem solving as one of the most difficult components of the math curriculum (Hudson & Miller, 2006). One instructional intervention in particular has shown evidence of improving the problem-solving performance of struggling students in middle school and high school. *Solve It!* (Montague, 2003) is a comprehensive routine that teaches students both the cognitive processes and the metacognitive strategies that successful problem solvers use while solving math word problems. The iterative process inherent to *Solve It!* is taught to students using the research-validated explicit instructional approach, whereby students learn through process modeling, mastery learning, and immediate and corrective feedback (Montague, 2003). Previous studies including single-subject designs as well as randomized group design have provided evidence for the effectiveness of *Solve It!* to improve the math problem-solving performance of participating students (Montague, 1992; Montague, Applegate, & Marquard, 1993; Montague & Bos, 1986; Montague, Enders, & Dietz, 2011).

A recent study examined the effectiveness of *Solve It!* when taught by general education math teachers to eighth-grade students of varying ability over the course of a school year. It is the purpose of the present study to determine whether the positive findings of that study can be replicated on a new, younger population of students (i.e., seventh-grade students). As such, we investigated the differential effects of the intervention on students of varying ability taught by general education math teachers in inclusive math classes. Further, we used Bayesian analyses to combine information from both studies (i.e., the eighth-grade study and the present seventh-grade study); this approach allowed us to synthesize the 2 years of intervention, producing more precise point estimates.

Mathematical Problem Solving

Mathematical problem solving involves multiple cognitive processes. Traditional models of problem solving identify two pri-

This article was published Online First December 2, 2013.

Marjorie Montague (deceased) and Jennifer Krawac, School of Education, University of Miami; Craig Enders, Department of Psychology, Arizona State University; Samantha Dietz, School of Education, University of Miami.

This research was supported by Institute for Education Sciences, U.S. Department of Education Grant R324A070206. Opinions expressed herein are those of the authors and do not represent the position of the U.S. Department of Education. The authors are grateful to the students as well as the teachers, administrators, and other school personnel in the Miami-Dade County Public Schools for their cooperation and support.

Correspondence concerning this article should be addressed to Craig Enders, Department of Psychology, Arizona State University, Box 871104, Tempe, AZ 85287-1104. E-mail: craig.enders@asu.edu

mary phases: problem representation and problem execution (Mayer, 1998; Polya, 1945/1986). These phases have been well supported across years of research and reflect the current problem-solving framework of the Program for International Student Assessment (PISA; Organization for Economic Co-operation and Development [OECD], 2013). Problem representation processes are needed to comprehend and integrate problem information, maintain mental images of the problem in working memory, and develop a viable solution path, with the objective of building "a coherent mental representation of the problem situation" (p. 126). Thus, problem representation requires students to translate linguistic and numerical information into verbal, graphic, symbolic, and quantitative representations that show how the problem information is related (Krawec, in press; Montague & Applegate, 1993; van Garderen & Montague, 2003). The translated information then helps generate appropriate mathematical equations or algorithms for problem solution. The second phase, problem execution, requires students to perform the appropriate calculations and to check for accuracy. Problem execution requires that students work forward and backward without resorting to a trial-and-error approach to problem solving.

Unsuccessful problem solvers often lack knowledge of (or fail to use) problem-solving processes, particularly those necessary for representing problems. Therefore, they need to be taught these processes explicitly and shown how to apply them when solving math word problems (e.g., Montague & Applegate, 1993). However, problem-solving instruction seldom addresses the processes in a comprehensive way. Rather, individual strategies are taught in isolation and are dispersed throughout the curriculum (e.g., Glencoe, 2004). To effectively teach students to solve math word problems, teachers must first understand the problem-solving processes and then provide appropriate comprehensive instruction with a variety of problem-solving activities (Kroll & Miller, 1993). *Solve It!*, the intervention in this school-based efficacy and replication study, identifies for teachers the processes and strategies proficient problem solvers use and provides a systematic instructional program to improve students' problem-solving performance (Montague, 2003).

The *Solve It!* Intervention

Solve It! is a cognitive strategy instructional approach that is based on scientifically demonstrated instructional principles and procedures (see Grouws, 1992; Swanson, Harris, & Graham, 2003). Cognitive strategy instruction is composed of the cognitive and metacognitive strategies that facilitate learning and improve performance, where the complexity of the strategies varies as a function of task difficulty. Though cognitive strategy instruction has been shown to improve the academic performance of students of varying disabilities (e.g., Coughlin & Montague, 2011), it is particularly effective for students with LD who do not know or use the types of strategies that support problem-solving proficiency (Graham & Harris, 2003; Montague & Dietz, 2009; Swanson, 2001). These students often use strategies despite proven ineffectiveness and do not generalize the use of strategies across academic domains (Swanson, 1993). Strategic learners, however, use a variety of strategies, depending on task demands, and use those strategies effectively and efficiently (Pressley, Borkowski, & Schneider, 1987).

Solve It! builds on and incorporates the cognitive processes that are integral to the development and application of declarative, procedural, and strategic knowledge of mathematics and to the ability to apply this knowledge when solving math word problems. The seven cognitive processes included in the routine reflect those validated by research (e.g., Berardi-Coletta, Dominowski, Buyer, & Rellinger, 1995; Krawec, in press; Montague, 1992; Montague, Enders, & Dietz, 2011). *Solve It!* places particular emphasis on teaching students how to comprehend, represent, and plan to solve mathematical problems. The ultimate goal of the program is to have students internalize the cognitive processes and metacognitive strategies so that they become automatic during problem solving. *Solve It!* incorporates the following cognitive processes and activities: reading (reading, rereading, identifying relevant/irrelevant information), paraphrasing (putting the problem into one's own words without changing the problem meaning), visualizing (transforming problem information to a representation that shows the relationships among problem parts), hypothesizing (setting up a plan to solve the problem by deciding on the type and order of operations), estimating (predicting the outcome based on the question/goal), computing (conducting the basic operations needed for solution), and checking (reviewing the accuracy of the process, procedures, and computation) (see Montague, Warger, & Morgan, 2000, for a more comprehensive description).

Cognitive processes may be described as online mental activities that are proactive in nature (the "to do" strategies), whereas metacognitive strategies require reflectivity and reactivity (the "what am I doing" and "what have I done" strategies). These emphasize self-awareness of cognitive knowledge; deployment of cognitive processes or strategies during problem solving; and control of strategies for purposes of regulating, evaluating, and monitoring performance (Berardi-Coletta et al., 1995). Problem solvers use metacognitive strategies to tell themselves what to do, ask themselves questions, recall what they know, detect and correct errors, and monitor performance. These strategies help problem solvers gain access to strategic knowledge, guide their application, and regulate use of strategies and overall performance as they solve problems. They can be used overtly (talking out loud or whispering to oneself) or covertly (silent self-talk). Metacognitive strategies include self-instruction, self-questioning, and self-monitoring. See Figure 1 for the *Solve It!* problem-solving routine.

Explicit instruction is the procedural approach used to teach students the *Solve It!* routine. Composed of research-based teaching strategies such as modeling, verbal rehearsal, and immediate feedback, explicit instruction is characterized by "structured, organized lessons; appropriate cues and prompts; guided and distributed practice; immediate and corrective feedback on learner performance; positive reinforcement; overlearning; and mastery" (Montague, Enders, & Dietz, 2011, p. 263). Explicit instruction, as compared with direct instruction, is more interactive as well as more flexible; it allows teachers to modify the routine based on students' specific needs (Montague, 2003; Montague et al., 2000).

Theoretical Underpinnings of *Solve It!*

The components of the *Solve It!* intervention are based on traditional problem-solving models (i.e., Mayer, 1985; Polya, 1945/1986) that have been incorporated into more current conceptualizations of problem solving. The OECD (2013) identified four

***Solve It!* - Math Problem Solving Processes and Strategies**

READ (for understanding)
Say: Read the problem. If I don't understand, read it again.
Ask: Have I read and understood the problem?
Check: For understanding as I solve the problem.
PARAPHRASE (your own words)
Say: Underline the important information. Put the problem in my own words.
Ask: Have I underlined the important information? What is the question? What am I looking for?
Check: That the information goes with the question.
VISUALIZE (a picture or a diagram)
Say: Make a drawing or a diagram. Show the relationships among the problem parts.
Ask: Does the picture fit the problem? Did I show the relationships?
Check: The picture against the problem information.
HYPOTHESIZE (a plan to solve the problem)
Say: Decide how many steps and operations are needed. Write the operation symbols (+, -, x, and /).
Ask: If I …, what will I get? If I …, then what do I need to do next? How many steps are needed?
Check: That the plan makes sense.
ESTIMATE (predict the answer)
Say: Round the numbers, do the problem in my head, and write the estimate.
Ask: Did I round up and down? Did I write the estimate?
Check: That I used the important information.
COMPUTE (do the arithmetic)
Say: Do the operations in the right order.
Ask: How does my answer compare with my estimate? Does my answer make sense? Are the decimals or money signs in the right places?
Check: That all the operations were done in the right order.
CHECK (make sure everything is right)
Say: Check the plan to make sure it is right. Check the computation.
Ask: Have I checked every step? Have I checked the computation? Is my answer right?
Check: That everything is right. If not, go back. Ask for help if I need it.

Figure 1. *Solve It!* math problem-solving processes and strategies. Adapted from *Solve It! A Mathematical Problem-Solving Instructional Program*, by M. Montague, 2003. Copyright 2003 by Exceptional Innovations.

processes involved in math problem solving in order to create an assessment for PISA. The four processes include exploring and understanding, representing and formulating, planning and executing, and monitoring and reflecting. The purpose and tasks associated with each process contributes to the overall goal of obtaining an accurate solution. The objective of exploring and understanding is to create mental representations of each piece of problem information and to identify and understand the relevant concepts. The second process, representing and formulating, requires that the solver select and organize the relevant information and then integrate it with the necessary background knowledge. The objective of the third process, planning and executing, is to determine the overall goal of the problem and devise a strategy to reach that goal using the requisite mathematical knowledge. Finally, the fourth process, monitoring and reflecting, requires that the solver evaluate progress throughout each stage and consider the assumptions underlying the solution as well as potential alternative solutions (OECD, 2013). Thus, the cognitive processes of *Solve It!* can be embedded into the PISA model to provide the "action" students need to use the knowledge at each step: exploring and understanding (read, paraphrase), representing and formulating (visualize), planning and executing (hypothesize, estimate, compute), and monitoring and reflecting (check).

Although the instructional content of *Solve It!* is based on the traditional problem-solving model, it alone is insufficient to effect change in students' problem-solving performance. For the intervention to be effective, both instructional content and instructional approach must be considered. As described above, *Solve It!* is structured on the principles of cognitive strategy instruction (CSI). On the basis of social development theory (Vygotsky, 1978), CSI provides the framework through which the research-based instructional content can be delivered. The active, engaged learning that is characteristic of CSI reflects the underlying roots of Vygotsky's theory of social development, where, through teacher facilitation, students are able to extend their understanding and demonstrate flexibility to successfully solve math word problems. Characteristics of *Solve It!* instruction, including teachers modeling the use of effective problem-solving strategies, verbal support and feedback, and reduced scaffolding over time, reflect the two main principles of social development theory: the more knowledgeable other and the zone of proximal development (Vygotsky, 1978). That is, students are able to internalize what they observe and practice during instruction, resulting in efficient and flexible use of effective problem-solving strategies (Wertsch, 1985).

Research Evidence of *Solve It!* Efficacy

Solve It! was first published as a curricular supplement in 2003 with over a decade of research supporting its effectiveness. Validation studies that used single-subject, quasi-experimental, and randomized control trial designs evidenced the success of the intervention in improving the problem-solving performance of students with LD. In 1986, Montague and Bos first tested the hypothesis that an eight-step cognitive strategy routine developed by the first author could improve the problem-solving performance of six high school students with LD who demonstrated adequate (i.e., fourth-grade) reading ability but difficulty in math. The single-subject multiple baseline across-participants design included baseline, treatment, and generalization/maintenance phases. Results showed that all students improved to criterion (at least 70% correct on four consecutive tests of math problem solving) and also demonstrated generalization of strategy use to more difficult problems. On the basis of this study, two adaptations to the cognitive routine were made: The eight cognitive processes were reduced to seven, and the metacognitive components of the strategy were made explicit.

In the second experiment, Montague (1992) investigated the effects of the intervention's two components—CSI and metacognitive strategy instruction (MSI)—on middle-school students' math problem solving. Six students participated in the multiple baseline across-participants design with two levels of treatment in order to conduct a componential analysis of the two types of instruction. During Treatment 1, the authors provided CSI alone to half the students and MSI alone to the other half and then added the complementary component during Treatment 2. Analysis of the treatment levels showed that students who received CSI then MSI reached mastery more quickly than those who had treatment in the reverse order. Overall, students' problem-solving performance improved on one-, two-, and three-step problems for all but the sixth graders. The authors inferred that the intervention may not be developmentally appropriate for younger students.

A quasi-experimental design was conducted by Montague, Enders, and Dietz (2009) in which seventh- and eighth-grade average-achieving students, low-achieving students, and students with LD received *Solve It!* instruction from their math teachers. Two months into the school year, 3 days of instruction using scripted lessons introduced the routine, requiring students to memorize the acronym (i.e., RPV-HECC) and apply the routine to word problems. Following the first week of instruction, one practice session per week was devoted to problem solving, in which students used the *Solve It!* routine to solve math word problems aligned to the district pacing guide. Results indicated that average-achieving students as well as low-achieving students and students with LD who received instruction ($n = 185$) made significantly greater progress in math problem solving over the course of the year than students in the comparison group ($n = 127$), as measured by curriculum-based measures (CBMs) of textbook-type problems ($d = .37$). Students in seventh and eighth grades had the same rate of growth, though, as expected, eighth-grade students' scores were higher.

Montague, Enders, and Dietz's (2011) most recent study included 24 middle schools matched on performance and socioeconomic status (SES). Schools from matched pairs were randomly assigned to treatment or comparison groups, and at each school, one eighth-grade math-certified teacher participated. A cluster-randomized design was used with a three-level model in which repeated measures were nested within students and students were nested within schools. Students were again provided 3 days of intensive instruction and then weekly practice sessions for the duration of the school year. Seven problem-solving CBMs were administered monthly to measure progress. Results indicated that students who received the intervention ($n = 319$) showed significantly greater growth in math problem solving over the course of the school year than students in the comparison group ($n = 460$) who received typical classroom instruction ($d = .91$). The intervention had the same impact on students across ability levels. Interestingly, though, by the end of the school year, students with LD in the treatment group significantly outperformed even the average-achieving students in the comparison group.

Thus, overall, findings related to the effectiveness of *Solve It!* have been validated across multiple studies. It was the purpose, then, of the present study to replicate the most recent study (i.e., Montague, Enders, & Dietz, 2011) with younger students. Further, we used a Bayesian analysis to synthesize results across 2 years of the intervention, incorporating prior information from the previous study of eighth graders. In the present study, *Solve It!* was investigated in the context of general education seventh-grade inclusive math classes with students with LD as well as low- and average-achieving students. On the basis of the previous efficacy study, we hypothesized that students receiving the intervention would outperform students in the comparison condition and that students with LD and low-achieving students would improve at a rate commensurate with their average-achieving peers. Two primary research questions guided this study.

1. What are the effects of *Solve It!* on math problem solving as measured by CBMs of problem solving for the intervention and comparison group students and subsets of students (i.e., students with LD, low-achieving students, average-achieving students)?

2. What are the effects of *Solve It!* on math and reading achievement as measured by the Florida Comprehensive Assessment Test (FCAT) math and reading tests for the intervention and comparison group students and subsets of students?

Method

Participants

Initially, 20 matched pairs of middle schools (i.e., 40 schools) were recruited from a possible 78 middle and K–8 schools in the Miami-Dade County Public Schools, the fourth largest school district in the nation serving approximately 340,000 students (9% White, 30% African American, 59% Hispanic, and 2% "other"; 60% districtwide qualify for the free/reduced lunch program). All middle schools and K–8 Centers were invited to participate. The sample consisted of the first 40 of the possible 78 schools whose principals agreed to participate in the study. Schools represented the range of FCAT performance levels and SES levels. School performance level was operationalized as the Florida Department of Education's assigned school grade (A, B, C, D, or F) based on FCAT performance. School-level SES was operationalized as the percentage of students who qualified for free or reduced lunch. Schools were paired by matching on the spectrum of FCAT performance level and SES to the extent possible. One school from each matched pair was randomly assigned to the intervention condition. A general education seventh-grade math teacher from each school was nominated by a school administrator to participate. The criteria for participation were that the teacher was "high quality," certified in mathematics education, teaching at least two class periods that included students with LD and low-achieving students, and, for the intervention teachers, willing to attend a 3-day *Solve It!* professional development workshop prior to the start of the school year. All teachers, both in treatment and in comparison groups, were provided compensation for their time commitment to the project; according to the university's internal review board, undue inducement (and thus a biased sample) was not a factor, as the amount and method of payment were deemed appropriate. Minimal attrition occurred before the intervention for the following reasons: one teacher did not attend the workshop, one teacher was reassigned by the principal, one teacher had only English as a Second Language classes, and three teachers decided not to participate. Thus, attrition reduced participating schools to 16 intervention schools and 18 comparison schools. No further attrition occurred for the remainder of the year; however, one intervention teacher and one comparison teacher left their respective schools midyear but were replaced by teachers who then consented to participate. We included these replacement teachers in the teacher demographics because they participated in the study. Thus, teacher demographics for 36 teachers are presented in Table 1.

All students from teachers' inclusion math, intensive math, general math, and pre-algebra class periods had equal opportunity to participate in the study (n = approximately 2,500). Participating classes at the schools ranged from two to six classes for each teacher. All students in the intervention schools were provided instruction, but data were collected only from students who returned consent forms ($n = 644$). Data were collected on all consented students in the comparison schools ($n = 415$). Participating students signed assent forms and returned consent forms signed by parents or legal guardians that described either the intervention or comparison condition. Attrition occurred in both groups throughout the year; our records showed that most often missing data resulted from a change in students' class schedules or a withdrawal from school due to transfer. Because our statistical analyses implemented a missing data-handling procedure that uses all available data, all consenting students were included in the student demographics table (see Table 2). The inclusion and intensive math classes typically included a higher percentage of students with LD and low-achieving students than other math classes. Group membership was operationalized as follows. Students with LD were district-identified[1] and had 2009 FCAT math score levels of 1 or 2 out of a possible 5. Low-achieving students also had 2009 FCAT math score levels of 1 or 2. In contrast, average-achieving students had 2009 FCAT math score levels of 3 or 4.

Table 1
Teacher Demographic Data

Variable	Intervention ($n = 17$) n (%)	Comparison ($n = 19$) n (%)
Gender		
Males	10 (59)	5 (26)
Females	7 (41)	14 (74)
Ethnicity		
White	1 (5)	2 (11)
Hispanic	11 (65)	8 (42)
African American	4 (25)	8 (42)
Missing	1 (5)	1 (5)
Level of education		
Bachelor's	9 (53)	9 (48)
Master's	6 (35)	8 (42)
Specialist/Doctoral	2 (12)	1 (5)
Missing		1 (5)
Years of teaching		
1–3	8 (47)	5 (26)
4–9	3 (18)	2 (11)
10 or more	6 (35)	12 (63)

Measures

CBMs. Seven CBMs of math problem solving were developed using 30 unique test items selected from the *Solve It!* manual (Montague, 2003) and calibrated using item response theory (IRT) methods (i.e., the Rasch model; Osterlind, 2000) to achieve equivalent difficulty across measures. Longitudinal analyses of CBMs are appropriate only when test scores from repeated administrations have been vertically equated to a common score scale (Montague, Penfield, Enders, & Huang, 2010). Consequently, the CBMs were first equated using the IRT procedure described in Montague et al. (2010). The internal consistency of the seven measures ranged from .71 to .79, very similar to that found in the previous study (Montague, Enders, & Dietz, 2011). Each of the measures consisted of 10 one-, two-, and three-step textbook-type math problems. The problems did not require specific or unique mathematical knowledge or concepts; rather, they required students to perform the four basic operations using whole numbers or decimals.

The FCAT. The FCAT's purpose is to assess student achievement of the Florida Sunshine State Standards. The assessment includes norm-referenced tests in reading and math, which allow for comparing the performance of Florida students with students across the nation. All students in Grades 3–10 take the statewide assessment test annually during March. The test measures a variety of skills, concepts, and applications. Passing the Grade 10 FCAT is a graduation requirement. FCAT reliability and validity are well established (Harcourt Asssessment, 2007). FCAT scale scores on the reading and math tests range between a hypothetical minimum of 100 and a maximum of 500; based on their scale scores, students are assigned an achievement level ranging from 1 to 5, where 1 is the lowest and 5 is the highest. Level 3 indicates performance on grade level.

[1] District eligibility criteria for placement in the learning disabilities program are as follows: (a) a disorder in one or more of the basic psychological processes including visual, auditory, or language processes; (b) academic achievement significantly below the student's level of intellectual functioning; (c) learning problems that are not due primarily to other disabling conditions; and (d) ineffectiveness of general educational alternatives in meeting the student's educational needs.

Table 2
Student Demographic Data

Variable	Intervention (n = 644) n (%)	Comparison (n = 415) n (%)
Ability level		
SLD	58 (9)	28 (7)
LAS	417 (65)	293 (70)
AAS	169 (26)	94 (23)
Gender		
Males	266 (41)	187 (45)
Females	378 (59)	228 (55)
Ethnicity		
White	38 (6)	17 (4)
Hispanic	405 (63)	278 (67)
African American	201 (31)	120 (29)
Free/reduced lunch		
Yes	501 (77)	345 (83)
No	143 (23)	70 (17)

Note. SLD = students with learning disabilities; LAS = low-achieving students; AAS = average-achieving students.

Procedure

Intervention teachers. The intervention teachers attended a 3-day professional development workshop in August prior to the start of school that included an overview of cognitive strategy instruction and the *Solve It!* approach, a description of the assessment tools and treatment fidelity checklists, demonstrations and modeling of the three lessons, practice using instructional procedures (e.g., verbal rehearsal), and a half-day break-out session for teachers to practice processing modeling (i.e., thinking aloud) while they solved problems. *Solve It!* includes a detailed instructional guide, scripted lessons, class charts, student cue cards, and practice problems. Intervention materials including 179 FCAT-style practice problems drawn from the district curriculum (i.e., the Glencoe 7th grade text, FCAT practice manuals, and previous FCAT math tests) were provided to teachers at the start of implementation (i.e., October); practice problems aligned to the district's scope and sequence were provided biweekly for the duration of the school year (see Florida Department of Education, 2006; fl.msmath2.net; Bailey et al., 2004). The teachers were given print copies as well as an electronic file of problems. Teaching scripts were provided for the initial three lessons and the practice sessions (a single template). The intervention began in October, continued across the school year, and consisted of 3 days of intensive instruction followed by weekly problem-solving practice sessions. Thus, students in the intervention group received *Solve It!* instruction embedded in the district curriculum, which included a mandated "pacing guide" that linked the Glencoe text to the Florida Sunshine State Standards.

Comparison teachers. Comparison teachers did not attend a workshop. They were instructed to proceed with "business as usual" (i.e., the district curriculum; Bailey et al., 2004). The first lesson in Chapter 1 of the Glencoe text is dedicated to a problem-solving strategy that teaches a four-step plan to solve a problem (i.e., Explore, Plan, Solve, Examine) and 12 "key concepts" and "problem-solving strategies" such as guess and check, look for a pattern, make an organized list, draw a diagram, act it out, and work backward (Bailey et al., 2004, pp. 6–7). Each subsequent chapter has a two-page supplementary section titled "Problem-Solving Strategy" that highlights and illustrates one of these strategies. Additionally, each chapter contains several "Practice: Word Problems" sections that contain six to eight problems requiring application of learned skills and concepts. Comparison group teachers were asked to focus on word problem solving during at least one class period per week. All students in both groups were allowed to use calculators during practice and testing sessions.

Progress monitoring. The CBMs were administered to the intervention group seven times, that is, prior to the intervention (baseline) and then approximately monthly for the remainder of the school year. The measures were administered four times to the comparison group, specifically, prior to the intervention (baseline) and then at the same time as the treatment group's third, fifth, and seventh administrations. The "missing data" were built into the research design as we decided a priori to test the comparison group on alternate occasions because the teachers and students might view seven administrations as excessive given that they were not receiving any direct benefit and no feedback. For the treatment group, beginning with the first CBM administration, the measure was scored, raw score data were entered into the database, and teachers were given printouts of results for all consented students. Following the third administration of the CBM, graphs for individual consented students in the intervention group were generated that included the first three data points. These graphs provided feedback to teachers and students regarding student progress over time and were redistributed following each subsequent CBM administration. By the end of the school year, each intervention student graph had seven data points. This procedure is part of the *Solve It!* instructional program.

Intervention group treatment fidelity. Observation checklists for the initial three lessons of the intervention and the practice sessions reflected teacher behaviors associated with preparation for each scripted lesson and implementation of the lesson. Each checklist for the three lessons contained 10, 15, and 15 items, respectively, that were scored as either "yes" or "no." That is, the behavior was either demonstrated or was not (e.g., "Did the teacher review the chart with the processes?"). Trained research assistants observed one intervention class period per day for the 3 days of initial instruction for each of the intervention teachers. Subsequent weekly practice sessions were observed by at least one research assistant; 10% of the practice sessions were observed by two assistants. The number of practice sessions varied by school due to individual school scheduling difficulties and district testing; practice sessions ranged from eight to 15 sessions, with a mean of 11 sessions. Level of treatment fidelity and interrater agreement were averaged across the observations separately for the initial three lessons and the practice sessions. Percentages were calculated by dividing number of agreements by agreements plus disagreements multiplied by 100. For the observations of the initial three lessons, fidelity averaged 97% (range = 90%–100%); 33% of the observations were conducted by two individuals and agreement averaged 99% (range = 94%–100%). For the practice sessions, fidelity averaged 93% (range = 77%–100%); 10% of the observations were conducted by two individuals and agreement averaged 99% (range = 88%–100%). The high levels of fidelity of implementation were expected as teachers had scripts for all lessons; further, at the end of observed lessons, teachers were given verbal feedback based on the observation checklist, which helped them to improve the fidelity of future lessons.

Comparison group observations. Comparison group teachers were observed at least once and were asked to teach a typical lesson on math word problem solving. Instructions to the graduate assistants were as follows: "Observe for 50 minutes of instructional time. Describe the content of the lesson. Describe the instructional methods, practices, strategies, and techniques. Be as complete in your description as possible." Observations ranged from 45 to 55 min. Observers sat unobtrusively at the back of the classroom and remained uninvolved. Due to scheduling problems and cancellations, only 25 observations were conducted. Of these, only 11 appeared to focus specifically on math problem solving. Teachers may have chosen not to follow instructions or may have assumed they were conducting lessons on problem solving if they used problems that were linguistically presented (e.g., Find 25% of 200). The observed math lessons were analyzed to determine instructional procedures implemented by the teachers in the comparison schools. All classes were teacher-directed rather than student-centered but involved high student participation. In most classes, teachers demonstrated solving a problem and then told students to work independently or in groups to practice solving several similar problems. They offered no further explanation on "the how and why" of solving math problems. All but one teacher requested student volunteers to show how the problem was solved or directed students to the front of the class to solve the problem on the board. However, all of the students at the board were "talked" through the process of solving the problem by the teacher who asked questions such as "What do we do first? What do we do next? Now, what do we do? Don't you think you should . . .? " Almost always, when a student made an error, the teacher immediately informed the student of the error and provided corrective feedback indicating where the error occurred and how it should be remedied. Often the teacher requested another student to give the student at the board the correct answer with such comments as "Jane, tell him what to do next" or "Who can finish it for him?" These were typical procedures across the comparison group classrooms.

Statistical Analysis

The data from the study were consistent with a three-level model in which repeated measures (Level 1) were nested within students (Level 2), and students were nested within schools (Level 3). Consequently, we used a series of three-level multilevel models (MLMs) to accommodate the nested data structure. Because the study used a cluster-randomized design where schools were randomly assigned to conditions, a Level 3 dummy variable that denoted participation in the intervention (0 = comparison school, 1 = intervention school) was the focal predictor in the MLMs. In particular, we were interested in assessing whether the two conditions differentially improved during the course of the school year (i.e., a Group × Time interaction).

We followed the following model-building process. First, we estimated a series of unconditional models (i.e., models with no predictors) that partitioned the score variation into three orthogonal sources: within-person variability, between-person variability, and between-school variability (note that it was not possible to separate classroom-level and school-level variation because a single teacher was responsible for all classes within a given school). The purpose of these initial analyses was to identify sources of variation that were pertinent for the subsequent analyses. Next, we examined a model that included a temporal predictor at Level 1 that expressed the timing of each measurement occasion in months since the first day of October. This parameterization defined the growth model intercept as the expected outcome score at the initial assessment. The purpose of this step was threefold: (a) determine the appropriate functional form of the growth trajectory, (b) evaluate between-student variability in growth trajectories, and (c) evaluate between-school variation in average growth rates. The analyses at this step prompted us to pursue linear growth models with random slopes at the student and school levels (Level 2 and Level 3, respectively).

The primary analyses involved two MLMs. We used the first model to evaluate the Group × Time interaction effect (i.e., the β_3 coefficient) while controlling for eight student-level covariates, as follows:

$$\begin{aligned} Y_{tij} = {} & \beta_0 + \beta_1(Months_{tij}) + \beta_2(Condition_j) \\ & + \beta_3(Condition_j)(Months_{tij}) + \beta_4(LowAch_{ij}) + \beta_5(LrnDis_{ij}) \\ & + \beta_6(FRLunch_{ij}) + \beta_7(Female_{ij}) + \beta_8(Black_{ij}) \\ & + \beta_9(Hispanic_{ij}) + \beta_{10}(ReadFCAT_{ij}) \\ & + \beta_{11}\left(MathFCAT_{ij}\right) + b_{0_{ij}} + b_{1_{ij}}\left(Months_{tij}\right) + u_{0_j} \\ & + u_{1_j}\left(Months_{tij}\right) + \varepsilon_{tij}, \end{aligned} \tag{1}$$

where β_0 is the average baseline problem-solving mean for the comparison schools, β_1 is the average monthly change for the comparison schools, β_2 is the baseline mean difference between the comparison and the intervention schools, and β_3 is the growth rate difference between the two groups (i.e., the Group × Time interaction). The covariates included two dummy codes representing a three-category ability variable (low-achieving vs. average-achieving students, LD vs. average-achieving students), a dummy code representing free or reduced lunch assistance, a gender dummy code, two dummy codes representing a three-category ethnicity variable (African American vs. White, and Hispanic vs. White), FCAT math achievement scores, and FCAT reading achievement scores (FCAT scores were standardized to have a mean of zero and standard deviation of one). Because the goal was to control for student- and school-level differences in the covariates, we grand mean centered these variables (Enders & Tofighi, 2007). Finally, the model included five residual terms: b_{0ij} is a Level 2 residual that captured baseline score differences among students, b_{1ij} is a Level 2 residual that reflected growth rate differences among students, u_{0j} is a Level 3 residual that captured baseline mean differences among schools, u_{1j} is a Level 3 residual that allowed the average change rates to differ across schools, and r_{ti} is Level 1 time-specific error.

In the second model, we added interaction terms to examine whether ability group (students with LD, low-achieving students, average-achieving students) moderated the impact of the intervention effect. The MLM is as follows:

$$\begin{aligned} Y_{tij} = {} & \beta_0 + \beta_1(Months_{tij}) + \beta_2(Condition_j) \\ & + \beta_3(Condition_j)(Months_{tij}) + \beta_4(LowAch_{ij}) \end{aligned}$$

$$+ \beta_5(LrnDis_{ij}) + \beta_6(LowAch_{ij})(Condition_j)$$
$$+ \beta_7(LrnDis_{ij})(Condition_j) + \beta_8(LowAch_{ij})(Months_{tij})$$
$$+ \beta_9(LrnDis_{ij})(Months_{tij})$$
$$+ \beta_{10}(LowAch_{ij})(Condition_j)(Months_{tij})$$
$$+ \beta_{11}(LrnDis_{ij})(Condition_j)(Months_{tij}) + \beta_{12}(FRLunch_{ij})$$
$$+ \beta_{13}(Female_{ij}) + \beta_{14}(Black_{ij}) + \beta_{15}(Hispanic_{ij})$$
$$+ \beta_{16}(ReadFCAT_{ij}) + \beta_{17}(MathFCAT_{ij}) + b_{0_{ij}}$$
$$+ b_{1_{ij}}(Months_{tij}) + u_{0_j} + u_{1_j}(Months_{tij}) + \varepsilon_{tij}. \quad (2)$$

This model allowed us to determine whether (a) the baseline difference between the intervention and comparison group varied as a function of ability level (β_6 and β_7), (b) the comparison group growth rate differed by ability level (β_8 and β_9), and (c) the intervention effect (i.e., the Group $\times$ Time interaction) differed across ability levels (β_{10} and β_{11}).

We used Bayesian estimation in the Mplus 7 software package (Muthén & Muthén, 1998–2012) for all analyses. Briefly, a Bayesian analysis combines information from two sources: a prior distribution and the data. Conceptually, the resulting point estimates are a weighted average of the estimates from these two sources. For the first set of analyses, we used so-called noninformative prior distributions that contribute little to no information to the analysis. We chose this approach for the initial analyses because the resulting point estimates are determined solely by the data and are equivalent in value to those from maximum likelihood estimation (Gelman, Carlin, Stern, & Rubin, 2009). For the second set of analyses, we used informative prior distributions. As noted previously, the data for the present article come from a multiyear intervention study (i.e., Montague, Enders, & Dietz, 2011). The Bayesian paradigm is ideally suited for our situation because it provides a principled mechanism for synthesizing results across 2 years of the intervention. To this end, we used the data from Montague et al. (2011) to formulate an informative prior distribution for each regression coefficient in the MLM. Procedurally, we used maximum likelihood estimation to fit the models in Equations 1 and 2 to the previous year's data, and we then specified each prior as a normal distribution with a mean and standard deviation equal to the previous year's estimate and standard error, respectively. To ensure comparability of the estimates across years, we adopted a common score scale for the CBM of math problem solving by assigning the IRT ability estimates to have a mean of 50 and a standard deviation of 10 at baseline; after standardizing FCAT scores, the predictor metrics were identical.

The Bayesian estimation has important advantages for our study. First, as already noted, the Bayesian framework provides a principled mechanism for synthesizing results from a multiyear intervention. The first set of Bayesian analyses (noninformative priors) provides estimates that are numerically comparable from a standard MLM with maximum likelihood estimation, and the second set of estimates (informative priors) is effectively a weighted average across the final 2 years of the intervention. Second, combining 2 years of data can dramatically improve precision, such that the resulting credible intervals (the Bayesian analog of confidence intervals) are narrower than those obtained by using only the current year's data. Third, the Markov Chain Monte Carlo algorithm (Gibbs sampler) used by the Bayesian estimator yields asymmetric credible intervals that do not rely on large-sample approximations. This is particularly advantageous for forming inferences about parameters that do not follow a normal sampling distribution (e.g., the variance estimates from an MLM). Finally, Bayesian estimation accommodates incomplete data on the outcome variable, such that every participant with at least one wave of outcome data contributes to the analysis (see Enders, 2010). As seen in supplemental Table 1, the proportion of complete data varied across waves, partly by design (comparison schools did not provide data at Waves 2, 4, and 6) and partly due to attrition. Bayesian estimation yields accurate parameter estimates under the so-called missing at random (MAR) mechanism, whereby missingness at a particular wave is explained by covariates and by problem-solving scores at other waves. Although there is no way to empirically verify or test the MAR mechanism, we believe that this assumption is quite sensible; our records indicated that the most common reasons for missing data were that students' class schedules were changed or they transferred out of the school.

Results

The purpose of the present study was to determine the effectiveness of a research-based problem-solving program to improve the mathematical problem solving of seventh-grade students of varying ability in inclusive general education math classes. To answer our research question, we used a series of three-level MLMs to accommodate the nested data structure, examining the effects of the intervention on students' performance on problem-solving CBMs as well as the math FCAT. Our analyses of the FCAT achievement test data revealed very little evidence supporting a Group $\times$ Time interaction effect. Although the effects were in the expected direction (i.e., the intervention group made slightly larger gains than the comparison group), they were not significantly different from zero. Consequently, we limit our presentation to the focal outcome variable, the problem-solving CBMs.

To begin, we evaluated the Group $\times$ Time interaction while controlling for a number of student-level covariates (see Equation 1). Table 3 gives the point estimates, posterior standard deviations (the Bayesian analog of standard errors), and 95% credible intervals from the analysis. Because we grand mean centered the covariates, the β_0 through β_3 coefficients describe the expected developmental trajectory for students who score at covariate means. As seen in the table, the comparison group had an average baseline score of $\tilde{\beta}_0 = 50.051$ and a monthly growth rate of $\tilde{\beta}_1 = .716$, whereas the intervention condition had a baseline mean of 49.504 ($\tilde{\beta}_0 + \tilde{\beta}_2$) and a monthly change rate of 1.323 ($\tilde{\beta}_1 + \tilde{\beta}_3$).[2] The 95% credible interval for β_2 contained zero, indicating that the group means were effectively identical at baseline. More importantly, the credible interval for β_3 (i.e., the Group $\times$ Time interaction or growth rate difference) did not contain zero (95% CI [.108, 1.098]), which suggests that the intervention group improved at a significantly higher rate than the comparison group.

To further illustrate the MLM parameter estimates, Figure 2 shows the model-implied developmental trajectories for the two

[2] We use the tilde accent to denote the posterior median as a point estimate.

Table 3
MLM Parameter Estimates: Group × Time Interaction Controlling for Student-Level Covariates

Parameter	Est.	Posterior *SD*	95% LCL	95% UCL
		Regression coefficients		
Comparison baseline (β_0)	50.051	0.749	48.496	51.455
Comparison growth (β_1)	0.716	0.183	0.354	1.077
Intervention baseline difference (β_2)	−0.547	0.989	−2.472	1.460
Intervention growth difference (β_3)	0.607	0.248	0.108	1.098
Low vs. average achieving (β_4)	−0.954	0.523	−1.980	0.069
LD vs. average achieving (β_5)	−1.583	0.806	−3.245	−0.074
Lunch assistance (β_6)	0.002	0.478	−0.949	0.943
Female (β_7)	0.167	0.369	−0.559	0.904
African American vs. White (β_8)	−2.624	0.915	−4.398	−0.738
Hispanic vs. White (β_9)	−0.653	0.828	−2.243	1.032
FCAT Reading (β_{10})	0.637	0.231	0.196	1.100
FCAT Math (β_{11})	4.655	0.267	4.135	5.186
		Variance estimates		
Student baseline (σ^2_{b0})	14.849	2.329	10.791	19.662
Student growth (σ^2_{b1})	0.142	0.071	0.026	0.302
Baseline-growth covariance (σ_{b0b1})	−0.012	0.322	−0.714	0.495
School mean baseline (σ^2_{u0})	4.606	2.429	1.73	11.2
School mean growth (σ^2_{u1})	0.334	0.139	0.166	0.679
Baseline-growth covariance (σ_{uou1})	−0.639	0.476	−1.859	−0.018
Residual (σ^2_{ε})	61.404	1.464	58.671	64.359

Note. MLM = multilevel modeling; Est. = estimate; LCL = lower credible limit; UCL = upper credible limit; LD = learning disabilities; FCAT = Florida Comprehensive Assessment Test.

conditions. The lines in the figure are simple slopes representing the expected performance of students scoring at the covariate means. The vertical separation of the lines at their rightmost points depicts the predicted mean difference at the end of the semester: $M_{\mathrm{I}} = 58.765$ versus $M_{\mathrm{C}} = 55.063$ for the intervention and comparison schools, respectively. Expressed relative to the average within-school standard deviation from the unconditional model ($SD = 6.044$), this mean difference corresponds to a standardized effect size of $d = .613$, which is slightly larger than Cohen's (1988) medium benchmark of $d = .50$ (i.e., controlling for covariates, the group means differed by roughly six tenths of a standard deviation at the end of the spring term).[3]

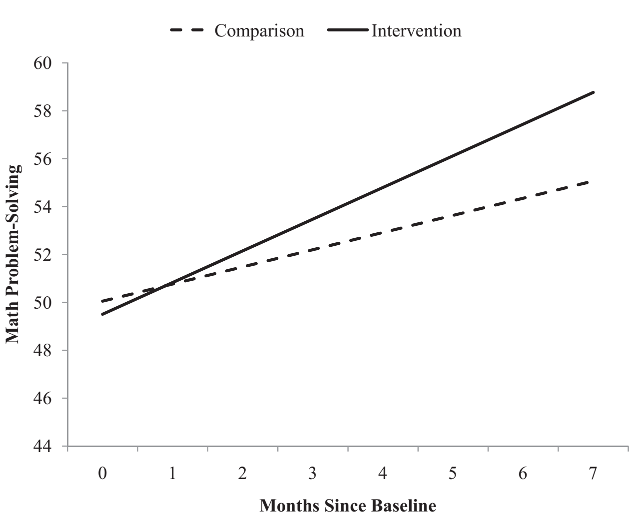

Figure 2. Model-implied growth trajectories (averages) from the curriculum-based measures model.

Having examined the intervention effect, we next estimated the MLM from Equation 2. The purpose of this model was to evaluate the moderating influence of ability group (LD, low-achieving students, average-achieving students) on the intervention effect. Specifically, the MLM allowed us to determine whether (a) the baseline difference between the intervention and comparison group varied as a function of ability level (β_6 and β_7), (b) the comparison group growth rate differed by ability level (β_8 and β_9), and (c) the Group × Time interaction varied by ability level (β_{10} and β_{11}). Supplemental Table 2 gives the point estimates, posterior standard deviations, and 95% credible intervals from the analysis. As seen in the table, the 95% credible intervals for β_4 and β_5 suggest that reliable ability group differences existed at baseline, such that low-achieving students and the students with LD scored lower than their average-achieving counterparts (the intervals do not contain zero). However, the credible intervals for the moderation coefficients ($\tilde{\beta}_6$ through $\tilde{\beta}_{11}$) contain zero, indicating that ability group had no material impact on baseline differences, developmental rates, or on the Group × Time intervention effect. For this reason, no further discussion of these results is warranted.

[3] As noted by Hedges (2007), cluster-randomized studies provide three different choices for the standard deviation in the denominator of the standardized mean difference effect size. We chose the pooled within-school standard deviation because it (a) is consistent with the corresponding standardizer from a single-level analysis of variance design and (b) yields an effect size that is applicable to a single-site study. We believe that the latter is particularly important because the resulting effect size provides a gauge of the treatment effect in a realistic situation in which the program is implemented in a single school.

Synthesizing Results Across Years

The Bayesian framework is compelling because it provides a mechanism for synthesizing results from the multiyear intervention. To this end, we used the data from the previous year's intervention (Montague, Enders, & Dietz, 2011) to formulate informative prior distributions for each regression coefficient in the MLM. Procedurally, we specified each prior as a normal distribution with a mean and standard deviation equal to the previous year's estimate and standard error, respectively. Specifying the priors in this manner yields point estimates that are effectively a weighted average of the estimates from the current data and the previous year. Because the variance estimates were not central to our major aims, we used standard noninformative priors for these parameters (i.e., the current year's data determined the variance estimates).

To begin, we estimated the MLM that included a Group × Time interaction and the student-level covariates (see Equation 1). Table 4 gives the point estimates, posterior standard deviations, and credible intervals from the analysis. As noted above, the point estimates can be viewed as a weighted average across the 2 years of the study. From a substantive perspective, the interpretation of the parameters is consistent with the previous analysis: $\tilde{\beta}_0 = 50.933$ and $\tilde{\beta}_1 = .340$ quantify the baseline mean and monthly growth rate, respectively, for the comparison group, and $\tilde{\beta}_2 = -1.521$ and $\tilde{\beta}_3 = .979$ capture the intervention group differences in these parameters. Importantly, the credible interval for β_3 (i.e., the Group × Time interaction or growth rate difference) did not contain zero (95% CI [.668, 1.298]), which suggests that the intervention group improved at a significantly higher rate than the comparison group. Note that combining the two data sources dramatically improved the precision of the estimates, as evidenced by a reduction in the posterior standard deviations (the Bayesian analog of standard errors) and a narrowing of the 95% credible intervals. This gain in precision (power) is a predictable consequence of integrating prior information about the parameters into the analysis.

To further illustrate the MLM parameter estimates, Figure 3 shows the model-implied developmental trajectories for the two conditions. The vertical separation of the lines at their rightmost points depicts the predicted mean difference at the end of the semester, $M_I = 58.645$ versus $M_C = 53.313$ for the intervention and comparison schools, respectively. Expressed relative to the average within-school standard deviation from the unconditional model ($SD = 6.044$), this mean difference corresponds to a standardized effect size of $d = .882$, which exceeds Cohen's (1988) benchmark for a large effect size.

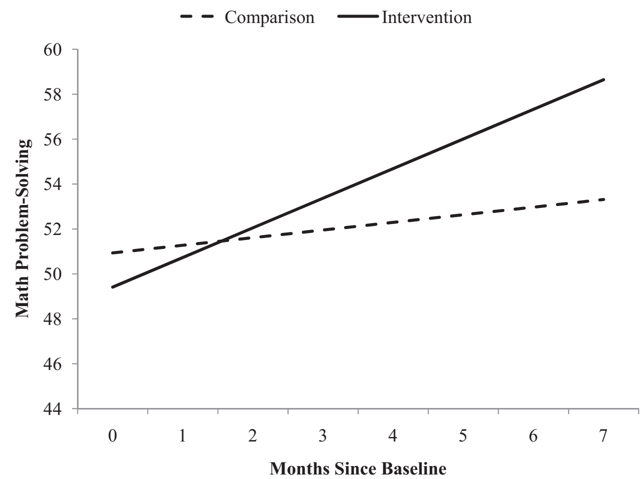

Figure 3. Model-implied growth trajectories (averages) from the curriculum-based measures model that synthesizes results across 2 years of the intervention.

Next, we estimated the MLM from Equation 2, again using informative prior distributions based on the previous year's regression coefficients. Supplemental Table 3 gives the point estimates, posterior standard deviations, and credible intervals from the analysis. Recall that the purpose of this model was to evaluate the moderating influence of ability group on the intervention effect. Consistent with the previous analysis, reliable ability group differences existed at baseline, such that low-achieving students and students with LD scored lower than their average-achieving coun-

Table 4
MLM Parameter Estimates Synthesizing Across Years: Group × Time Interaction Controlling for Student-Level Covariates

Parameter	Est.	Posterior *SD*	95% LCL	95% UCL
Comparison baseline (β_0)	50.933	0.410	50.125	51.741
Comparison growth (β_1)	0.340	0.110	0.119	0.560
Intervention baseline difference (β_2)	−1.521	0.578	−2.697	−0.407
Intervention growth difference (β_3)	0.979	0.157	0.668	1.298
Low vs. average achieving (β_4)	−1.256	0.396	−2.034	−0.473
LD vs. average achieving (β_5)	−2.227	0.633	−3.550	−1.069
Lunch assistance (β_6)	0.185	0.368	−0.561	0.912
Female (β_7)	−0.164	0.258	−0.672	0.356
African American vs. White (β_8)	−2.336	0.507	−3.327	−1.325
Hispanic vs. White (β_9)	−0.629	0.408	−1.410	0.182
FCAT Reading (β_{10})	0.762	0.173	0.430	1.111
FCAT Math (β_{11})	4.489	0.217	4.060	4.913

Note. MLM = multilevel modeling; Est. = estimate; LCL = lower credible limit; UCL = upper credible limit; LD = learning disabilities; FCAT = Florida Comprehensive Assessment Test.

terparts (the 95% credible intervals do not contain zero). More importantly, the analysis suggests that the intervention effect was somewhat stronger for low-achieving students than for average-achieving students, $\tilde{\beta}_{10}$ = .484, 95% CI [.204, .735]. The growth rate difference for students with LD in the intervention condition was also positive, $\tilde{\beta}_{11}$ = .362, 95% CI [−.062, .727], but we cannot conclude that their monthly gains exceeded average-achieving students because the 95% credible interval contains zero.

Discussion

The goal of the present study was to improve the problem solving of seventh-grade students of varying abilities through teacher implementation of the research-based instructional program *Solve It!*. This study was a replication of a previous study that included older students (Montague, Enders, & Dietz, 2011). Progress monitoring through CBMs of math problem solving over the school year showed that, like the eighth-grade students in the previous study, seventh-grade students who received the intervention embedded in the district curriculum showed significantly greater growth in math problem solving than the comparison group students who received only the district curriculum. Although the difference in growth rate between groups on the math test of the FCAT was not statistically significant, a small effect on growth rate from 2009 to 2010 was evident for the intervention group only. Moreover, the intervention appeared to have a similar effect on growth rates for students with LD, low-achieving students, and average-achieving students. Overall, for students across ability groups, it had no material impact on baseline differences, developmental rates, or on the magnitude of the intervention effect, although, as one would expect, average-achieving students performed statistically significantly better initially than students with LD on all criterion measures and low-achieving students on the FCAT only and continued that advantage. The difference in growth rates on math problem solving between the intervention and comparison groups also was practically significant with a difference between groups of more than one half of a standard deviation (d = .613), a medium effect size, for the final administration of the CBMs. Importantly, after synthesizing the results of the two studies (i.e., the present study and Montague, Enders, & Dietz, 2011) to improve the precision of the estimates, the intervention effect (d = .882) exceeded Cohen's (1988) benchmark for a large effect size. Thus, the findings generally support the efficacy of *Solve It!* as an intervention to improve math problem solving for middle-school students.

Several of these findings are noteworthy. First, though the intervention did not statistically improve the performance of the intervention students on the math test of the FCAT, the high-stakes state assessment test that is a graduation requirement in the state of Florida, a small effect was found for the intervention group, largely attributable to the increase in scores for the students with LD and the low-achieving students. No difference in growth rate or in growth effect between intervention and comparison groups was evident on the FCAT reading test, which seems to further validate the improvement in mathematics on the state assessment. Again, it should be noted that there were no slope differences within a given ability group between the treatment and comparison conditions; however, a small effect was evident for the intervention group but not the comparison group. The FCAT math test reports are global in the sense that they produce a composite score; that is, they do not break out scale scores that would provide insight into performance on various strands of math skills, concepts, and applications tested (e.g., computation, applied problem solving, measurement, geometry). Thus, it is not surprising that the intervention, which addressed only one of the several components of the FCAT, did not effect significant change in performance. The statistically nonsignificant improvement, however, still translates into practically important gains in performance on a test that determines grade promotion, eligibility for certain courses, and even graduation (Florida Department of Education, 2012).

Second, the very large effect (d = .882) of the intervention on the problem-solving performance of students of varying ability supports the use of *Solve It!*, an instructional intervention originally designed for students with LD, in inclusive settings composed of varying levels of ability. These findings are particularly timely as schools are moving toward full inclusion of students with LD in general education classes (U.S. Department of Education National Center for Education Statistics, 2010). According to Glazzard (2011), teachers in inclusive classrooms reported that finding appropriate instructional materials that meet all students' needs is one of the biggest challenges of inclusion. Indeed, over a decade of research on teacher perceptions and attitudes toward inclusion reported this same concern (Buell, Hallam, & Gamel-McCormick, 1999; Schumm & Vaughn, 1995). Because this intervention improved the performance of students across ability levels, *Solve It!* may be the solution for embedding problem-solving instruction into the curriculum of inclusive classrooms. Further, Yell and Walker (2010) recommended that schools use "mathematics programming that research has shown to be effective" (p. 135) and also that schools "adopt and use research-based progress monitoring systems to collect data on student performance" (p. 135). The *Solve It!* intervention reflects the characteristics of instruction as recommended by Yell and Walker, with the present study further validating the research-based effectiveness of *Solve It!*.

Finally, it is important to note that, after synthesizing results from the previous study, the intervention effect was stronger for low-achieving students than for average-achieving students, and, although not statistically significant, students with LD also showed a positive growth rate difference compared with their average-achieving peers. Research suggests that in contrast to reading, where children with lower initial scores tend to make faster progress over time than children with higher initial scores, in math, children with both low- and high-initial scores seem to progress at the same rate, thus maintaining the discrepancy apparent early in school into elementary, middle, and high school (Montague, Enders, & Castro, 2006; Montague, Enders, Cavendish, & Castro, 2011; Morgan, 2009). That is, children with higher initial scores maintain higher performance, whereas those with lower scores maintain low scores across time. Interestingly, the present study only partially follows that pattern. In our sample, the lowest performing students (particularly the low-achieving students) in the treatment group closed the performance gap between themselves and their average-achieving peers, who also improved, but at a slower rate of growth.

These findings suggest that incorporating specialized, evidence-based instruction in math problem solving into the curriculum may have a positive impact on students' math achievement trajectories in the middle-school years. In the present study, *Solve It!* instruction effectively improved the problem-solving performance of low-achieving students, average-achieving students, and students with LD. Better performance in math problem solving should improve students' overall math performance, particularly as new state standards are being implemented that highlight problem-solving proficiency (Common Core State Standards Initiative, 2012).

Limitations of the Study

Three limitations to the study should be noted. First, professional development was not provided to the comparison group teachers. Consequently, it is possible that the intervention effects may have been due, in part, to the training the intervention teachers received in teaching problem solving, irrespective of content, which occurred prior to the implementation of *Solve It!*. However, the intention of the study was to compare *Solve It!* instruction, which was embedded in the district curriculum, with the district curriculum only (i.e., typical math instruction or "business as usual"). Second, intervention teachers provided, to the extent possible, weekly problem-solving practice sessions, whereas the comparison group teachers were asked to focus on problem solving for one class period per week. The textbooks (i.e., the curriculum) provided problem-solving instruction and included multiple practice problems within each chapter. However, it remains possible that intervention students had more opportunities to engage in problem solving, which may have affected research outcomes. Finally, professional development on the *Solve It!* routine for treatment teachers occurred in August, but implementation of the intervention did not begin until mid-October. Until the pretest was administered, no materials (i.e., scripted lessons, cue cards, word problems, etc.) were provided to teachers, and, further, they were told to teach the district curriculum only without any *Solve It!* instruction until mid-October when the intervention began; however, the possibility exists that teachers incorporated some aspect of *Solve It!* into their lessons prior to the intervention.

Implications and Future Research Directions

This study provides insight into the National Council of Teachers of Mathematics (NCTM) goal of linking research to practice and effectively implementing research-based instructional procedures in general education classrooms. First, the general education math teachers expressed a desire for more professional development related to the instructional components of *Solve It!* (e.g., ongoing monitoring, feedback, modeling) in order to successfully implement this intervention. Second, they noted the pressures associated with adhering to the district curriculum and preparing students for the FCAT, the state high-stakes test. These concerns are important to consider because many teachers will not implement an intervention unless it can be incorporated seamlessly into the curriculum and does not impose further burdens on their primary teaching responsibilities. For these reasons, we embedded *Solve It!* into the curriculum to the extent possible by using problems drawn directly from the district-mandated textbooks. Problems linked to state standards and aligned to the district pacing guide were then provided to teachers weekly for the duration of the practice sessions. Teachers thus perceived the intervention as complementary and embedded rather than as supplementary. A third concern relating to NCTM's goal involves the teaching methodology and pace of instruction. Our general education math teachers needed extensive practice in cognitive or process modeling, as it was different from their typical instructional approach (i.e., didactic instruction and worked examples). Therefore, cognitive modeling was emphasized during the professional development workshop, and the research team modeled problem solving as needed in the classrooms over the course of the project. Additionally, some teachers struggled with how long students took initially to solve problems, as some students took as long as 20 min to solve one problem initially. This procedure also is counter to instruction in most general education math classrooms. However, as implementation progresses, the time factor ceases to be a concern as modeling shifts its focus to developing the ability to evaluate problems, think flexibly, and become more efficient. In other words, modeling begins to emphasize how to "size up" a problem with respect to its difficulty and then effectively and efficiently apply the *Solve It!* routine. By midyear, students, rather than the teacher, should be modeling all of the practice problems. Also, by midyear, students should be able to solve and model three or four problems during a class period. By the end of the year, students should be able to complete 10 one-, two-, and three-step problems and score within the "zone" (i.e., 7, 8, 9, or 10 problems correct) within a class period. For our sample, by the end of the school year, 62.1% of the students in the intervention group had scored within the zone.

Overall, the purpose of the present study was to conduct a randomized clinical trial in general education math classes to investigate the efficacy of an intervention to improve math problem solving for middle-school students. Although the results were positive and promising, several concerns were noted that have implications for "linking research and practice." Future studies should consider what is needed to maintain evidence-based practices in general education classrooms. Improving math problem solving for students, particularly those students who perform poorly, remains an essential need in our schools.

References

Bailey, R., Day, R., Frey, P., Howard, A. C., Hutchens, D. T., McClain, K., . . . Willard, T. (2004). *Mathematics applications and concepts*. New York, NY: Glencoe/McGraw-Hill.

Berardi-Coletta, B., Dominowski, R. L., Buyer, L. S., & Rellinger, E. R. (1995). Metacognition and problem solving: A process-oriented approach. *Journal of Educational Psychology, 21,* 205–223.

Buell, M. J., Hallam, R., & Gamel-McCormick, M. (1999). A survey of general and special education teachers' perceptions and inservice needs concerning inclusion. *International Journal of Disability, Development and Education, 46,* 143–156. doi:10.1080/103491299100597

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum.

Common Core State Standards Initiative. (2012). *Mathematics: Standards for mathematical practice*. Retrieved from http://www.corestandards.org/Math/Practice

Coughlin, J., & Montague, M. (2011). The effects of cognitive strategy instruction on the mathematical problem solving of adolescents with spina bifida. *Journal of Special Education, 45,* 171–183. doi:10.1177/0022466910363913

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevelmodels: A new look at an old issue. *Psychological Methods, 12,* 121–138. doi:10.1037/1082-989X.12.2.121

Florida Department of Education. (2006, May 23). *2006 FCAT: Reading and mathematics grades 3–10*. Retrieved from http://fcat.fldoe.org/mediapacket/2006/pdf/FCAT_06_media_complete.pdf

Florida Department of Education. (2012). *Understanding FCAT 2.0 reports: Spring 2012*. Retrieved from http://fcat.fldoe.org/fcat2/pdf/s12uf2r.pdf

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2009). *Bayesian data analysis*. New York, NY: Chapman & Hall/CRC.

Glazzard, J. (2011). Perceptions of the barriers to effective inclusion in one primary school: Voices of teachers and teaching assistants. *British Journal of Learning Support, 26,* 56–63. doi:10.1111/j.1467-9604.2011.01478.x

Graham, S., & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 323–334). New York, NY: Guilford Press.

Grouws, D. (1992). *Handbook of research on mathematics teaching and learning*. New York, NY: MacMillan.

Harcourt Assessment. (2007). *Florida Comprehensive Assessment Test: Reading and Mathematics technical report*. San Antonio, TX: Author.

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32,* 341–370.

Hudson, P., & Miller, S. P. (2006). *Designing and implementing mathematics instruction for students with diverse learning needs*. New York, NY: Pearson.

Krawec, J. (in press). Problem representation and mathematical problem solving of students of varying ability. *Journal of Learning Disabilities*.

Kroll, D. L., & Miller, T. (1993). Insights from research on mathematical problem solving in the middle grades. In D. T. Owens (Ed.), *Research ideas for the classroom: Middle grades mathematics* (pp. 58–77). New York, NY: MacMillan.

Mayer, R. E. (1985). Mathematical ability. In R. J. Sternberg (Ed.), *Human abilities: Information processing approach* (pp. 127–150). San Francisco, CA: Freeman.

Mayer, R. E. (1998). Cognitive, metcognitive, and motivational aspects of problem solving. *Instructional Science, 26,* 49–63. doi:10.1023/A:1003088013286

Montague, M. (1992). The effects of cognitive and metacognitive strategy instruction on mathematical problem solving of middle school students with learning disabilities. *Journal of Learning Disabilities, 25,* 230–248. doi:10.1177/002221949202500404

Montague, M. (2003). *Solve It! A mathematical problem-solving instructional program*. Reston, VA: Exceptional Innovations.

Montague, M., & Applegate, B. (1993). Mathematical problem-solving characteristics of middle school students with learning disabilities. *The Journal of Special Education, 27,* 175–201. doi:10.1177/002246699302700203

Montague, M., Applegate, B., & Marquard, K. (1993). Cognitive strategy instruction and mathematical problem-solving performance of students with learning disabilities. *Learning Disabilities Research and Practice, 8,* 223–232.

Montague, M., & Bos, C. (1986). The effect of cognitive strategy training on verbal math problem solving performance of learning disabled adolescents. *Journal of Learning Disabilities, 19,* 26–33. doi:10.1177/002221948601900107

Montague, M., & Dietz, S. (2009). Evaluating the evidence base for cognitive strategy instruction and mathematical problem solving. *Exceptional Children, 75,* 285–302.

Montague, M., Enders, C., & Castro, M. (2006). Academic and behavioral outcomes for students at risk for emotional and behavioral disorders. *Behavioral Disorders, 30,* 87–96.

Montague, M., Enders, C., Cavendish, W., & Castro, M. (2011). Academic and behavioral trajectories for at-risk adolescents in urban schools. *Behavioral Disorders, 36,* 141–156.

Montague, M., Enders, C., & Dietz, S. (2009). [Effects of cognitive strategy instruction on the math problem solving of middle school students]. Unpublished raw data.

Montague, M., Enders, C., & Dietz, S. (2011). Effects of cognitive strategy instruction on math problem solving of middle school students with learning disabilities. *Learning Disability Quarterly, 34,* 262–272. doi:10.1177/0731948711421762

Montague, M., Penfield, R. D., Enders, C., & Huang, J. (2010). Curriculum-based measurement of math problem solving: A methodology and rationale for establishing equivalence of scores. *Journal of School Psychology, 48,* 39–52. doi:10.1016/j.jsp.2009.08.002

Montague, M., Warger, C., & Morgan, H. (2000). Solve It! Strategy instruction to improve mathematical problem solving. *Learning Disabilities Research and Practice, 15,* 110–116. doi:10.1207/SLDRP1502_7

Morgan, P. (2009). *The trajectories of kindergarten children's mathematical achievement*. Paper presented at the annual IES project directors' meeting, Washington, DC.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Organization for Economic Co-operation and Development. (2013). *PISA 2012. Assessment and analytical framework: Mathematics, reading, science, problem solving, and financial literacy*. Retrieved from. doi:10.1787/9789264190511-en

Osterlind, S. J. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Polya, G. (1986). *How to solve it: A new aspect of mathematical method*. Princeton, NJ: Princeton University Press. (Original work published 1945)

Pressley, M., Borkowski, J. G., & Schneider, W. (1987). Cognitive strategies: Good strategy users coordinate metacognition and knowledge. In R. Vasta & G. Whitehurst (Eds.), *Annals of child development* (Vol. 5, pp. 89–129). New York, NY: JAI Press.

Schumm, J., & Vaughn, S. (1995). Getting ready for inclusion: Is the stage set? *Learning Disabilities Research and Practice, 10,* 169–179. Retrieved from http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1540-5826

Swanson, H. L. (1993). Principles and procedures in strategy use. In L. Meltzer (Ed.), *Strategy assessment and instruction for students with learning disabilities* (pp. 61–92). Austin, TX: ProO-Ed.

Swanson, H. L. (2001). Research on interventions for adolescents with learning disabilities: A meta-analysis of outcomes related to higher-order processing. *The Elementary School Journal, 101,* 331–348.

Swanson, H. L., Harris, K. R., & Graham, S. (Eds.). (2003). *Handbook of learning disabilities*. New York, NY: Guilford Press.

U.S. Department of Education National Center for Education Statistics. (2010). *The Condition of Education 2010*. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010028

van Garderen, D., & Montague, M. (2003). Visual-spatial representations and mathematical problem solving. *Learning Disabilities Research and Practice, 18,* 246–254. doi:10.1111/1540-5826.00079

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wertsch, J. V. (1985). *Vygotsky and the social formation of the mind*. Cambridge, MA: Harvard University Press.

Yell, M. L., & Walker, D. W. (2010). The legal basis of response to intervention: Analysis and implications. *Exceptionality, 18,* 124–137. doi:10.1080/09362835.2010.491741

Received January 11, 2011
Revision received September 18, 2013
Accepted October 10, 2013 ■