# Does Comparing Solution Methods Facilitate Conceptual and Procedural Knowledge? An Experimental Study on Learning to Solve Equations

Bethany Rittle-Johnson
Vanderbilt University

Jon R. Star
Harvard University

Encouraging students to share and compare solution methods is a key component of reform efforts in mathematics, and comparison is emerging as a fundamental learning mechanism. To experimentally evaluate the effects of comparison for mathematics learning, the authors randomly assigned 70 seventh-grade students to learn about algebra equation solving by either (a) comparing and contrasting alternative solution methods or (b) reflecting on the same solution methods one at a time. At posttest, students in the compare group had made greater gains in procedural knowledge and flexibility and comparable gains in conceptual knowledge. These findings suggest potential mechanisms behind the benefits of comparing contrasting solutions and ways to support effective comparison in the classroom.

*Keywords:* learning processes, conceptual (declarative) knowledge, procedural knowledge, mathematics education, mathematics concepts

Current educational reforms in mathematics advocate that the teacher act more as a facilitator, encouraging students to share and compare their own thinking and problem-solving methods with other students (Hiebert & Carpenter, 1992; National Council of Teachers of Mathematics, 1991, 2000). Despite an abundance of descriptive research suggesting the promise of this approach, experimental studies that demonstrate its benefits are largely absent. In this study, we experimentally evaluated a potentially pivotal component of this instructional approach that is supported by basic research in cognitive science: the value of students comparing multiple examples. We used a unique design that allowed for random assignment to condition within intact classrooms to increase both internal and external validity. Seventh-grade students learned about algebra equation solving by either (a) comparing and contrasting alternative solution methods or (b) reflecting on the alternative solution methods one at a time. The findings have theoretical implications for when and why contrasting examples

facilitate learning and practical implications for how to support effective comparison in the classroom.

## Comparing Alternative Solution Methods

For at least the past 20 years, a central tenet of reform pedagogy in mathematics has been that students benefit from comparing, reflecting on, and discussing multiple solution methods (Silver, Ghousseini, Gosen, Charalambous, & Strawhun, 2005). Case studies of expert mathematics teachers emphasize the importance of students actively comparing solution methods (Ball, 1993; Fraivillig, Murphy, & Fuson, 1999; Huffred-Ackles, Fuson, & Sherin Gamoran, 2004; Lampert, 1990; Silver et al., 2005). For example, having students share and compare multiple methods was a distinguishing factor of the most skilled teacher in a sample of 12 first-grade teachers implementing a new reform mathematics curriculum (Fraivillig et al., 1999). Furthermore, teachers in high-performing countries such as Japan and Hong Kong often have students produce and discuss multiple solution methods (Stigler & Hiebert, 1999). This emphasis on sharing and comparing solution methods was formalized in the National Council of Teachers of Mathematics *Standards* (1989, 2000).

Although these and other studies provide evidence that sharing and comparing solution methods is an important feature of expert mathematics teaching, existing studies do not directly link this teaching practice to measured student outcomes. We could find no studies that assessed the causal influence of comparing contrasting methods on student learning gains in mathematics.

There is, however, a robust literature in cognitive science that provides empirical support for the benefits of comparing contrasting examples for learning in other domains (mostly in laboratory settings; Gentner, Loewenstein, & Thompson, 2003; Kurtz, Miao, & Gentner, 2001; Loewenstein & Gentner, 2001; Namy & Gentner, 2002; Oakes & Ribar, 2005; Schwartz & Bransford, 1998). For example, college students who were prompted to compare two business cases by reflecting on their similarities were much more likely to transfer the solution strategy to a new case than were

students who read and reflected on the cases independently (Gentner et al., 2003). Thus, identifying similarities and differences in multiple examples may be a critical and fundamental pathway to flexible, transferable knowledge. However, this research has not been done in mathematics, with K–12 students, or in classroom settings.

In the current study, we extended the existing educational and cognitive science research on contrasting examples by using a randomized design to evaluate whether comparing solution methods promoted greater learning in mathematics than studying methods one at a time. We focused on three critical components of mathematical competence: procedural knowledge, procedural flexibility, and conceptual knowledge (Hiebert, 1986; Kilpatrick, Swafford, & Findell, 2001). Procedural knowledge is the ability to execute action sequences to solve problems, including the ability to adapt known procedures to novel problems (the latter ability is sometimes labeled *transfer*; Rittle-Johnson, Siegler, & Alibali, 2001). Procedural flexibility incorporates knowledge of multiple ways to solve problems and when to use them (Kilpatrick et al., 2001; Star, 2005, 2007) and is an important component of mathematical competence (Beishuizen, van Putten, & van Mulken, 1997; Blöte, Van der Burg, & Klein, 2001; Dowker, 1992; Star & Seifert, 2006). Finally, conceptual knowledge is "an integrated and functional grasp of mathematical ideas" (Kilpatrick et al., 2001, p. 118). This knowledge is flexible and not tied to specific problem types and is therefore generalizable (although it may not be verbalizable). Overall, we hypothesized that comparing solution methods would lead to greater procedural knowledge, flexibility, and conceptual knowledge.

## Importance of Algebra

We evaluated the effectiveness of comparing multiple solution methods for learning a pivotal component of mathematics—algebra. Historically, algebra has represented students' first sustained exposure to the abstraction and symbolism that makes mathematics powerful (Kieran, 1992). Regrettably, students' difficulties in algebra have been well documented in national and international assessments (Blume & Heckman, 1997; Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999). One component of algebra, linear equation solving, is considered a basic skill by many in mathematics education and is recommended as a curriculum focal point for Grade 7 by the National Council of Teachers of Mathematics (Ballheim, 1999; National Council of Teachers of Mathematics, 2006). When introduced, the methods used to solve equations are among the longest and most complex to which students have been exposed. Current mathematics curricula typically do not focus sufficiently on flexible and meaningful solving of equations (Kieran, 1992).

Examples of different equation solving methods as applied to several types of linear equations are shown in Table 1. In the center column of the table are conventional and commonly taught methods for solving linear equations that apply to most equations. In the right-most column of the table are nonconventional methods that treat expressions such as $(x + 1)$ as composite variables (with the exception of the fourth row). The first three methods in the right-most column are arguably shortcuts—they are more efficient because they involve fewer steps and fewer computations; thus they may be executed faster and with fewer errors. Procedural flexibility requires that students understand important problem

Table 1
*Alternative Solution Methods for Four Types of Equations*

| Equation type[a] | Sample solution via conventional method | Sample solution via nonconventional method |
|---|---|---|
| $a(x + b) = c$<br>Divide composite | $3(x + 1) = 15$<br>$3x + 3 = 15$<br>$3x = 12$<br>$x = 4$ | $3(x + 1) = 15$<br>$x + 1 = 5$<br>$x = 4$ |
| $a(x + b) + d(x + b) = c$<br>Combine composite | $2(x + 1) + 3(x + 1) = 10$<br>$2x + 2 + 3x + 3 = 10$<br>$5x + 5 = 10$<br>$5x = 5$<br>$x = 1$ | $2(x + 1) + 3(x + 1) = 10$<br>$5(x + 1) = 10$<br>$x + 1 = 2$<br>$x = 1$ |
| $a(x + b) = d(x + b) + c$<br>Subtract composite | $7(x - 2) = 3(x - 2) + 16$<br>$7x - 14 = 3x - 6 + 16$<br>$7x - 14 = 3x + 10$<br>$4x - 14 = 10$<br>$4x = 24$<br>$x = 6$ | $7(x - 2) = 3(x - 2) + 16$<br>$4(x - 2) = 16$<br>$x - 2 = 4$<br>$x = 6$ |
| $a(x + b) + dx + e = f(gx + h) + ix + c$<br>Conventional | $4(x - 2) + 2x + 10 = 2(3x + 1) + 4x + 8$<br>$4x - 8 + 2x + 10 = 6x + 2 + 4x + 8$<br>$6x + 2 = 10x + 10$<br>$2 = 4x + 10$<br>$-8 = 4x$<br>$-2 = x$ | $4(x - 2) + 2x + 10 = 2(3x + 1) + 4x + 8$<br>$4(x - 2) + 2x + 2 = 2(3x + 1) + 4x$<br>$4x - 8 + 2x + 2 = 6x + 2 + 4x$<br>$6x - 6 = 10x + 2$<br>$-6 = 4x + 2$<br>$-8 = 4x$<br>$-2 = x$ |

[a] All $x$s stand for variables; all other letters were replaced with numbers.

features and identify the most efficient method for solving a given problem.

## Current Study

We compared learning from comparing multiple solutions (*compare* group) to learning from studying sequentially presented solutions (*sequential*, or *control*, group) for seventh-grade students learning to solve multistep linear equations such as $3(x + 1) = 15$ and $5(y + 2) = 2(y + 2) + 12$. Students in both conditions studied worked examples of hypothetical students' solution methods and answered questions about the methods with a partner.

Three features of our study design merit a brief justification. First, we chose to provide students with worked examples because doing so insured exposure to multiple methods for all students and facilitated side-by-side comparison of these methods for students in the compare condition. Many studies have shown that students from elementary school to university—both in the laboratory and in the classroom—learn more efficiently and deeply if they study worked examples paired with practice problems rather than solving the equivalent problems on their own (see Atkinson, Derry, Renkl, & Wortham, 2000, for a review). Second, we chose to have students work with a partner because past research indicates that students who collaborate with a partner tend to learn more than those who work alone (e.g., Johnson & Johnson, 1994; Webb, 1991), and teaching students to generate conceptual explanations for their partners improves their own learning (Fuchs et al., 1997). And third, we chose to prompt students to generate explanations when studying worked examples because there is a great deal of evidence that doing so leads to greater learning, as compared to cases when students are not asked to provide explanations (e.g., Bielaczyc, Pirolli, & Brown, 1995; Chi, de Leeuw, Chiu, & La-Vancher, 1994).

Pairs of students were randomly assigned to condition and completed the intervention during 2 days of partner work within their intact mathematics classrooms. We hypothesized that students in the compare group would show greater improvements from pretest to posttest on three outcome measures—(a) procedural knowledge (particularly transfer), (b) procedural flexibility, and (c) conceptual knowledge—than students in the sequential group. We expected these differences to emerge as a result of students making more explicit comparisons between methods, which should highlight the accuracy and efficiency of multiple solution methods and facilitate adoption of nonstandard methods.

## Method

### Participants

All 70 seventh-grade students at a selective, private, urban school participated (36 girls, 34 boys). There were four seventh-grade mathematics classes at the school; two regular and two advanced, with 36 students in the advanced classes and 14–20 students per class. Students' mean age was 12.8 years (range: 11.7–13.8 years); 81% were Caucasian, 10% African American, 3% Asian, 3% Indian, and 3% Middle Eastern; and approximately 10% received financial aid. About three quarters of students typically enter the school in kindergarten, in which admission is based on basic school readiness. On the Educational Records Bureau's

*Comprehensive Testing Program* (2004), seventh graders at this school on average score above the 80th percentile nationally on the quantitative and mathematics sections.

There was one seventh-grade mathematics teacher, who was a Caucasian male with 5 years of teaching experience and an undergraduate degree in mathematics. In all four classes, the teacher used the *Passport to Algebra and Geometry* text (Larson, Boswell, Kanold, & Stiff, 1999), and students in the advanced classes went more in depth on a subset of lessons. In previous lessons, students had learned about the distributive property, simplifying expressions, and solving one-step and simple two-step equations. The teacher indicated that he sometimes encouraged use of multiple solution methods. Human subjects' approval and all relevant consents (from the head of the middle school, the teacher, the parents, and the students) were obtained before the study began.

### Design

We used a pretest–intervention–posttest design. For the intervention, students were randomly paired with another student in their class (without regard to student characteristics such as gender), and then pairs of students were randomly assigned to condition, with approximately equal numbers of pairs in each condition within each class. Pairs in the compare condition ($n = 18$ pairs, 8 of them mixed-gender pairs) studied sets of two worked examples for the same problem and answered questions encouraging comparison of the two examples. Pairs in the sequential condition ($n = 17$ pairs, 8 of them mixed-gender pairs) studied the same two worked examples on two isomorphic problems and answered questions encouraging reflection on a single example. Students also solved practice problems and received mini-lectures from the teacher during the intervention.

### Materials

*Intervention.* Four types of equations were used during the intervention, as shown in Table 1. The worked examples illustrated two different solution methods for each problem, typically a conventional method for solving the equations and a shortcut method for solving the equations that relied on treating subexpressions as a composite variable and reduced the number of computations and steps needed to solve the equation (see Table 1). The exception was the fourth problem type (see Table 1, fourth row), where the second method shown in the worked example was a *less* efficient method that involved subtracting a term from both sides as a first step. This fourth problem type was included so that the conventional method was sometimes the most efficient method.

Packets of worked examples were created for each condition. In the compare packets, there were 12 equations (three instances of each of the four types), with each equation solved in two different ways, presented side by side on the same page for a total of 24 worked examples. Each step was labeled using one of four step labels (distribute, combine, add/subtract on both, multiply/divide on both). On some examples, students were asked to label some of the steps to encourage active processing of the examples. At the bottom of the page were two questions prompting students to compare and contrast the two worked examples. A sample page from the packet is shown in Panel A of Figure 1. There was a

### A. Compare Condition

| Mandy's Solution: | | Erica's Solution: | |
|---|---|---|---|
| $5(y + 1) = 3(y + 1) + 8$ | | $5(y + 1) = 3(y + 1) + 8$ | |
| $5y + 5 = 3y + 3 + 8$ | *Distribute* | $2(y + 1) = 8$ | *Subtract on Both* |
| $5y + 5 = 3y + 11$ | *Combine* | $y + 1 = 4$ | *Divide on Both* |
| $2y + 5 = 11$ | *Subtract on Both* | $y = 3$ | *Subtract on Both* |
| $2y = 6$ | *Subtract on Both* | | |
| $y = 3$ | *Divide on Both* | | |

    1. Mandy and Erica solved the problem differently, but they got the same answer. Why?

    2. Why might you choose to use Erica's way?

### B. Sequential Condition

**Mandy's Solution:**

| | |
|---|---|
| $5(y + 1) = 3(y + 1) + 8$ | |
| $5y + 5 = 3y + 3 + 8$ | *Distribute* |
| $5y + 5 = 3y + 11$ | *Combine* |
| $2y + 5 = 11$ | *Subtract on Both* |
| $2y = 6$ | *Subtract on Both* |
| $y = 3$ | *Divide on Both* |

_____

    1. Would you choose to use Mandy's way to solve problems like this? Why or why not?
-----NEXT PAGE-----

**Erica's Solution:**

| | |
|---|---|
| $10(x + 3) = 6(x + 3) + 16$ | |
| $4(x + 3) = 16$ | *Subtract on Both* |
| $x + 3 = 4$ | *Divide on Both* |
| $x = 1$ | *Subtract on Both* |

_____

    1. Check Erica's solution by substituting her answer into the equation. Did Erica get the right answer?

*Figure 1.* Sample pages from intervention packet for (A) compare and (B) sequential conditions.

separate packet for each of the two days of partner work; the first two problem types in Table 1 were presented in the first packet, and the third and fourth problem types in Table 1 were presented in a second packet.

In the sequential packets, there were 24 equations, the 12 equations from the compare condition and an isomorphic equation for each that was identical in form and varied only in the particular numbers. The same solution methods were presented as in the compare condition, but each worked example was presented on a separate sheet. Thus, exposure to multiple solution methods was equivalent across the two conditions. As in the compare condition, steps were labeled or students needed to fill in the appropriate label. At the bottom of each page was one question prompting students to reflect on that solution. The number of reflection questions (24) was the same across the two conditions. A pair of sample pages from the packet is shown in Panel B of Figure 1.

There was also a packet of 12 practice problems. The problems were isomorphic to the equations used in the worked examples, and the same practice problems were used for both conditions. Three brief homework assignments were developed, primarily using problems in the students' regular textbook, and homework was the same for both conditions.

*Assessment.* The same assessment was used as an individual pretest and posttest. It was designed to assess procedural knowledge, flexibility, and conceptual knowledge. Sample items of each knowledge type are shown in Table 2. The procedural knowledge items were four familiar equations (one of each type presented during the intervention) and four novel, transfer equations (e.g., a problem that included three terms within parentheses). There were six flexibility items designed to tap three components of flexibility—the abilities to generate, recognize, and evaluate multiple solution methods for the same problem. There were six conceptual

Table 2
*Sample Items for Assessing Procedural Knowledge, Flexibility, and Conceptual Knowledge*

| Problem type | Sample items | Scoring |
|---|---|---|
| **Procedural knowledge** | | |
| Familiar ($n = 4$) | $-1/4\ (x - 3) = 10$ <br> $5(y - 12) = 3(y - 12) + 20$ | 1 pt for each correct answer. |
| Transfer ($n = 4$) | $0.25\ (t + 3) = 0.5$ <br> $-3(x + 5 + 3x) - 5(x + 5 + 3x) = 24$ | 1 pt for each correct answer. |
| **Flexibility** | | |
| Generating multiple methods ($n = 2$) | Solve this equation in two different ways: $4(x + 2) = 12$ | 1 pt if two different, correct solutions. |
| Recognize multiple methods ($n = 2$) | For the equation $2(x + 1) + 4 = 12$, identify all possible steps that could be done next. (4 choices) | 1 pt for each correct choice. |
| Evaluate nonconventional methods ($n = 2$) | $3(x + 2) = 12$ <br> $x + 2 = 4$ | |
| | a. What step did the student use to get from the first line to the second line? | a. 1 pt if correctly identify step. |
| | b. Do you think that this way of starting this problem is (a) a very good way; (b) OK to do, but not a very good way; (c) not OK to do? | b. 2 pts for choice a, 1 pt for choice b. |
| | c. Explain your reasoning. | c. 3 pts if justify and say quicker/easier; 2 pts for quicker/easier; 1 pt if don't reject, but prefer alternative. |
| **Conceptual knowledge** ($n = 6$) | 1. If $m$ is a positive number, which of these is equivalent to (the same as) $m + m + m + m$? (Responses are: $4m$; $m^4$; $4(m + 1)$; $m + 4$.) | 1 pt for selecting $4m$. |
| | 2. Here are two equations: <br> $213x + 476 = 984$ <br> $213x + 476 + 4 = 984 + 4$ | |
| | a. Without solving either equation, what can you say about the answers to these equations? (Responses are: both answers are the same; both answers are different; I can't tell without doing the math.) | a. 1 pt for selecting "both answers are the same." |
| | b. Explain your reasoning. | b. 2 pts if justify that same thing done to both sides doesn't change value of $x$, 1 pt if note that 4s cancel out. |

*Note.* pt = point.

knowledge items designed to tap students' verbal and nonverbal knowledge of algebra concepts, such as maintaining equivalence and the meaning of variables. Both the assessment and the intervention packets are available from Bethany Rittle-Johnson upon request.

## Procedure

All data collection occurred within students' intact mathematics classes over four consecutive 45-min classroom periods (the experimental manipulation occurred on 2 of these days). The instruction replaced the students' regular instruction on solving multistep linear equations and occurred immediately after regular instruction on solving basic two-step linear equations. On Day 1, students first completed the pretest. Students were given 30 min to complete the pretest, including 16 min to complete the eight procedural knowledge items. Some time pressure was included for the procedural knowledge items to encourage students to use efficient solution methods. After students completed the pretest, the teacher presented an equation to the class, $2(y - 3) + 5y = 22$, and asked them to attempt to solve the problem on their own. Then he presented a brief, scripted lesson to the entire class on solving the equation using the distributive property, as this was students' first formal exposure to solving equations with parentheses. The

teacher noted that there were multiple ways to solve the problem, but he only presented one way. All students were given the same brief homework assignment.

On Day 2, the experimental manipulation was introduced (recall that students from both conditions worked within the same classroom). Students sat with their partner and were given the appropriate packet of worked examples and practice problems for the day (covering the first two equation types in Table 1). Students were instructed to alternate between studying two worked examples with their partner and solving a practice problem on their own. Interleaving practice problems with worked examples helps students to monitor their understanding of solving the problems (Atkinson et al., 2000). When studying the worked examples, they were instructed to describe each solution to their partner and answer the accompanying questions first verbally and then in writing. Partners had a single worked-example packet and were encouraged to take turns writing their responses. The written explanation served to push students to summarize their ideas and come to a consensus. Student pairs' verbal interactions were taperecorded to provide supplemental qualitative data.

Students had their own practice problem packet and were asked to solve each problem on their own, compare answers with their

partner, and ask for help if the answers were not the same. The classroom teacher and two members of the project team (one of whom was typically Bethany Rittle-Johnson) circulated through the class, answering student questions and making sure that students were complying with directions. The teacher and project members provided help implementing steps (e.g., how to divide both sides by 1/4), but not choosing solution steps or answering reflection questions. Student pairs worked at their own pace and were not expected to complete the packet. They were also not aware that different pairs of students were studying different packets. All students were given the same homework assignment at the end of the class period.

On Day 3, the third and fourth equation types shown in Table 1 were covered. The teacher first provided a brief, scripted, whole-class lesson on the conventional method for solving the challenge problem from the previous day's homework and emphasized that there was more than one way to solve the problem. The problem, $15t = 5(2t - 7)$, was students' first formal exposure to equations with variables on both sides. Then, students sat with their same partner and worked on the second packet of worked examples and practice problems, under the same circumstances as Day 2. All students were given the same homework assignment.

On Day 4, the teacher provided a brief summary lesson (approximately 10 min) to the entire class. In this scripted summary, he emphasized that (a) there is more than one way to solve an equation, and any way is OK as long as you always keep the two sides of the equation equal; and (b) some ways to solve an equation are better than others because they are easier for you or because they make it less likely that you will make a mistake. Finally, students were given 30 min to complete the posttest, which was identical in content and administration to the pretest.

There were no differences in lessons or packet materials for the regular and advanced classes. However, students in the advanced classes completed more of the partner packets. On average, advanced students studied 22 of the available 24 worked examples and solved 11 of the available 12 practice problems, compared to regular students' studying 19 worked examples and completing 9 practice problems.

To ensure fidelity of treatment, all whole-class lessons were scripted, and a member of the project team followed along with the script during each lesson and verified that each key idea was presented and that additional information was not added. During the partner work, help guidelines (as described above) were followed by the teacher and project team members. Observations by the first author during the intervention and when reviewing transcripts of partner work for three pairs indicated that these guidelines were followed, that the two conditions did not receive different levels of help, and that the students did not notice that different pairs were working on different packets during classwork (homework assignments did not differ by condition).

### Coding

*Assessment.* The eight problems on the pretest and posttest procedural knowledge assessment were scored for accuracy of the answer. Cronbach's alpha was .70 at pretest and .58 at posttest. This reliability was sufficient for making group comparisons; for group size of at least 25, the probability of expected reversals of two groups (i.e., the probability that the lower-scoring group would surpass the higher-scoring group if they were tested again with the same test) is less than .05 when Cronbach's alpha is .50 (Thorndike, 1997). If anything, lower reliabilities lead to underestimates of effects (Thorndike, 1997). The test–retest correlation was .60, and interscorer agreement, computed on 20% of responses by two independent scorers, was 100%.

In addition to scoring accuracy, students' solution methods were coded into one of four categories—conventional method, shortcut method, other method, and blank. For this coding, computational errors were ignored. The conventional method was defined as first distributing, then combining like terms (if possible), then adding/subtracting from both sides, and finally dividing/multiplying on both sides (e.g., as shown in Table 1). The demonstrated shortcut method was defined as using one of the shortcut steps demonstrated in the worked examples (e.g., divide composite, combine composite, and subtract composite; see Table 1). In each shortcut, rather than distributing first, students treated expressions of the form $(x + a)$ as a variable and divided, combined like terms, or subtracted from both sides. This code was not relevant on two problems (the 4th learning problem type and its associated transfer problem). All other attempted solution methods were coded as "other," and they included use of other nonconventional methods, methods that violated mathematical principles, and incomplete methods that were too ambiguous to code as either conventional or shortcut use. The fourth category was for blanks (e.g., students did not attempt to solve the problem). Interrater agreement for solution method was calculated for 20% of the sample at pretest and posttest by an independent coder, and exact agreement was 100% for use of the demonstrated shortcut and 90% for use of the conventional method ([number of agreements/number of items] × 100).

The flexibility assessment had three components (see Table 2 for scoring details). The percentage of possible points on each component was calculated, and the three percentages were averaged to yield an overall flexibility score. Cronbach's alpha was .72 at pretest and .78 at posttest, the test–retest correlation was .57, and interscorer agreement on 20% of responses was 93%.

On the conceptual knowledge assessment, students received one point for correctly answering each of the six objective questions correctly (see Table 2 for scoring criteria). In addition, students explained their reasoning on two items, and these explanations were scored on a 2-point scale (interscorer agreement on 20% of explanations was 87%). These explanation scores were added to students' conceptual knowledge totals. Thus, a conceptual knowledge score was calculated as a percentage of possible points. Cronbach's alpha was .61 at pretest and .59 at posttest, and the test–retest correlation was .65.

For each assessment, we calculated students' *gain score* as posttest minus pretest. We opted to analyze gain scores, rather than posttest scores (with pretest score as a covariate), because either analysis method is equally acceptable for two-wave data, and gain scores are more straightforward to interpret.

*Intervention.* Recall that students solved practice problems during the intervention and that we tallied how many problems each student completed (students found the correct solution before moving on, so accuracy was not scored). We also coded whether students used the demonstrated shortcut method to solve the problems, and interrater reliability on 20% of the sample was 100%.

Student pairs also provided written explanations during the intervention. Two coding schemes were developed to code these explanations, and these will be discussed in the Results section. Exact agreement on presence of each explanation type, conducted by two raters on 20% of the sample, ranged from 90% to 100%.

## Data Analysis

Because students worked with a partner for the intervention, we calculated intraclass correlations to test for nonindependence in partner scores (Grawitch & Munz, 2004; Kenny, Kashy, Mannetti, Pierro, & Livi, 2002). Indeed, partners' gain scores were often modestly related (*r*s in the .2 range), violating assumption of nonindependence in traditional analysis of variance models. Following the recommendations of Kenny et al. (2002), we used multilevel modeling, incorporating their actor-partner interdependence model (see http://davidakenny.net/dyad.htm for a tutorial and details on implementing this approach in SPSS). As indicated by Kenny et al. (2002), we specified the use of restricted maximum likelihood estimation and compound symmetry for the variance–covariance structure in the models. The significance tests used the Satterthwaite (1946) approximation to estimate the degrees of freedom, which generally results in fractional degrees of freedom (see Kenny et al., 2002).

Our model had two levels—the individual level and the dyad level. To incorporate the actor-partner interdependence model into the model, we included the partner's scores as predictors in the first-stage (individual-level) analyses (Kenny et al., 2002). In other words, both a person's own pretest scores and their partner's pretest scores were used as predictors of the individual's outcomes in the first-stage analyses. Effect of experimental condition was tested in the second-stage (dyad-level) analyses. Because students were tracked by ability, and given that students of higher ability may learn at a faster rate or respond differently to instructional manipulations, the effect of ability group was also included in the second-stage analyses. We did not expect ability group to interact with experimental condition, and preliminary analyses indicated that it did not, so the interaction term was not included in the final models.

One student was absent on the day of the posttest. Statisticians strongly recommend the use of imputation, rather than the traditional procedure of omitting participants with missing data, because it leads to more precise and unbiased conclusions (Peugh & Enders, 2004; Schafer & Graham, 2002). When the data are missing at random, and less than 5% of the data is missing, as in this case, simulation studies indicate that single imputation leads to the same conclusions as when there are no missing data (Barzi & Woodward, 2004; Harrell, 2001). The student's missing posttest scores were imputed by regression from nonmissing values using the IMPUTE procedure of Stata 9. The single imputation model included all the independent and dependent variables that were included in subsequent analyses on accuracy scores, as described below.

To estimate the practical significance of differences between conditions, we computed effect sizes (Cohen's *d*) as the difference in gain scores between conditions divided by the pooled standard deviation of the gain scores.

## Results

We first overview students' knowledge at pretest. Next, we report the effect of condition on gains in students' knowledge from pretest to posttest. Finally, we examine the effects of the manipulation during the intervention; in particular, we report on solution methods and explanation quality during the intervention.

## Pretest Knowledge

Recall that our intervention occurred after students had completed lessons on solving basic one- and two-step equations. Thus, at pretest, students had some algebra knowledge. As shown in Table 3, students solved one or two of the equations correctly and had some success on the measures of flexibility and conceptual knowledge. When solving the equations, they most often used the conventional method and left a fair number of the problems incomplete or blank (see Table 4). Procedural knowledge correlated with both conceptual knowledge, $r(68) = 0.39$, $p = .001$, and flexibility, $r(68) = 0.32$, $p = .01$, but flexibility and conceptual knowledge were not related, $r(68) = 0.11$, $p = .38$.

At pretest, there were no significant differences between conditions on the procedural knowledge, flexibility, or conceptual knowledge measures, $F(1, 68) = 1.39$, 2.85, and 0.94, respectively (see Table 3), nor did the two conditions differ in their classroom grades for the previous grading period, $F(1, 68) = 0.01$. Male and female students did not differ in success on the pretest measures.

## Knowledge Gains From Pretest to Posttest

Students in the compare condition were expected to make greater gains from pretest to posttest in procedural knowledge,

Table 3
*Student Performance by Condition*

| Condition | Pretest | | Posttest | | Gain | | |
|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | Cohen's *d* |
| Compare | | | | | | | |
| Procedural | 16.7 | 22.5 | 50.7 | 23.7 | 34.0 | 20.6 | .53 |
| Flexibility | 32.3 | 23.7 | 69.1 | 16.6 | 36.7 | 16.7 | .38 |
| Conceptual | 46.5 | 16.6 | 59.4 | 21.5 | 12.9 | 18.9 | −.14 |
| Sequential | | | | | | | |
| Procedural | 22.8 | 23.7 | 46.4 | 23.7 | 23.6 | 19.0 | |
| Flexibility | 26.2 | 15.0 | 55.9 | 25.2 | 29.7 | 19.7 | |
| Conceptual | 51.6 | 21.3 | 67.1 | 19.8 | 15.5 | 16.8 | |

Table 4
*Solution Method by Condition (Proportion of Trials)*

| Solution method | Pretest | | Posttest | |
|---|---|---|---|---|
| | Compare | Sequential | Compare | Sequential |
| Conventional | 0.46 | 0.43 | 0.61 | 0.66† |
| Demonstrated shortcut | 0.00 | 0.00 | 0.17 | 0.10* |
| Other | 0.38 | 0.40 | 0.17 | 0.19 |
| Blank | 0.16 | 0.17 | 0.05 | 0.05 |

*Note.* Differences between conditions were significant with multilevel modeling as marked.
† $p$ (1, 31.6) = .06.   * $p$ (1, 30.8) < .05.

procedural flexibility (measured via solution strategy use and via an independent measure), and conceptual knowledge. Multilevel modeling was used to evaluate the effect of condition on gain scores on each of the measures. In each analysis, the individual's pretest score and his or her partner's pretest score on that measure were included as predictors at the first level (these scores were standardized to facilitate interpretation of effects). Condition and ability group were included as predictors at the second (dyad) level. In the initial models, students' gender and whether the student was in a same-gender or mixed-gender pair were included as predictor variables. However, neither were significant predictors in any of the models (all $ps$ > .15), so neither variable was included in the final models. See Table 5 for a summary of the final models. Note that when posttest scores, rather than gain scores, were used as the dependent variable, the parameter estimates and $t$ values for the second-level predictors (condition and ability group) were identical. The only substantive difference between the posttest score and gain score models was that an individual's pretest score positively predicted posttest scores, as expected.

*Procedural knowledge.* Students in the compare condition made greater gains in procedural knowledge (see Table 3 and

Figure 2). There was a main effect for condition and for the individual's pretest score (see Table 5). Students in the compare condition gained 10 additional percentage points compared to those in the sequential condition ($d$ = .53). In addition, lower pretest knowledge was associated with higher gain. Those with more to learn learned more.

*Flexibility: Solution methods on procedural knowledge items.* We expected students in the compare condition to become more flexible, as well as more accurate, problem solvers. One measure of flexibility was using the demonstrated shortcuts rather than the conventional method when appropriate. Indeed, as shown in Table 4, students in the compare condition were more likely to use the demonstrated shortcuts at posttest, $t(30.8)$ = 2.06, $p$ = .048, $d$ = .34, than students in the sequential condition. This increased shortcut use seemed to partially account for (i.e., mediate) the benefits of the compare condition for accuracy. When frequency of using a shortcut method at posttest was included in the model of procedural knowledge gain reported above, shortcut use positively predicted accuracy gain, $t(61.6)$ = 2.41, $p$ = .019, and condition no longer did, $t(32.0)$ = 1.33, $p$ = .19.

*Flexibility: Independent measure.* Students in the compare condition also made greater gains on the independent measures of flexibility (see Table 3 and Figure 3). There were main effects for condition, ability group, and the individual's pretest score (see Table 5). Students in the compare condition gained an additional 7 percentage points ($d$ = .39). Students in the advanced classes also made greater gains, as did students with lower pretest knowledge. The items asking students to generate multiple solutions to the same problem may have been biased in favor of the compare condition, given that these students saw (although did not generate) multiple solutions to the same problem. As a result of this possible bias, we repeated the flexibility analyses, excluding these items from the scores. There continued to be a main effect for condition, $t(31.5)$ = 2.45, $p$ = .020, $d$ = .14, ability group, $t(32.0)$ = 3.95, $p$ < .001, and individual's pretest score, $t(64.7)$ = −8.71, $p$ < .001.

Table 5
*Multilevel Modeling Results for Three Student Learning Outcomes*

| Parameter | Coefficient | SE | df | t |
|---|---|---|---|---|
| Procedural knowledge gain | | | | |
| Intercept | .20 | .04 | 31.00 | 5.27*** |
| Condition (compare) | .08 | .04 | 31.00 | 2.12* |
| Ability group (advanced) | .07 | .05 | 31.00 | 1.44 |
| Own pretest score | −.10 | .03 | 64.25 | −4.04*** |
| Partner's pretest score | .04 | .03 | 64.24 | 1.63 |
| Flexibility gain | | | | |
| Intercept | .18 | .03 | 31.00 | 5.55*** |
| Condition (compare) | .10 | .04 | 31.00 | 2.78** |
| Ability group (advanced) | .20 | .04 | 31.00 | 5.41*** |
| Own pretest score | −.07 | .02 | 64.86 | −3.44*** |
| Partner's pretest score | .01 | .02 | 64.87 | 0.52 |
| Conceptual knowledge gain | | | | |
| Intercept | .14 | .04 | 31.00 | 3.41** |
| Condition (compare) | −.04 | .04 | 31.00 | −0.91 |
| Ability group (advanced) | .04 | .06 | 31.00 | 0.61 |
| Own pretest score | −.10 | .02 | 49.85 | −4.22*** |
| Partner's pretest score | .03 | .02 | 49.85 | 1.59 |

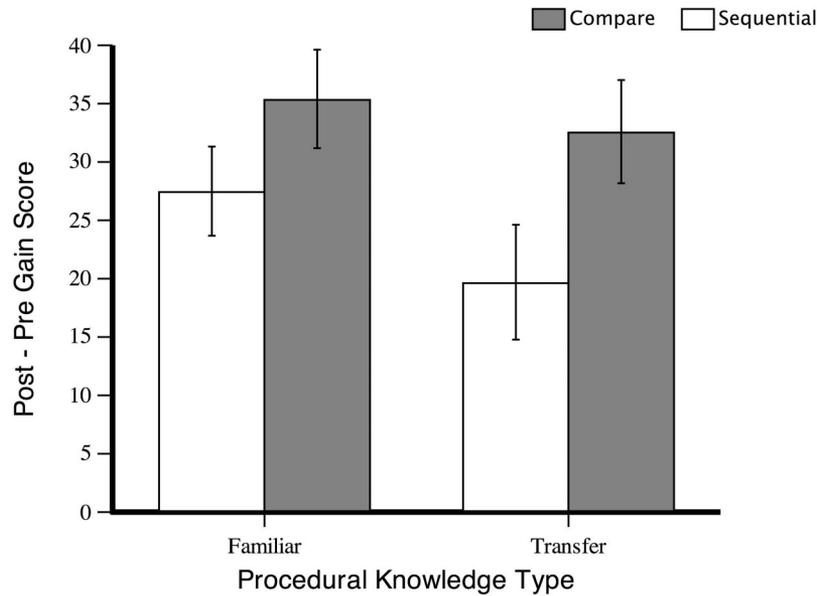* $p$ < .05.   ** $p$ < .01.   *** $p$ < .001.

*Figure 2.* Procedural knowledge gain for familiar and transfer problems, by condition (error bars are SE).

*Conceptual knowledge.* Students in the compare and sequential conditions did not differ in their conceptual knowledge gain (see Table 3). The only significant predictor of gain was the individual's pretest score (see Table 5). Although there was no difference between conditions, students across conditions did show improvements in conceptual knowledge from pretest to posttest ($M = 49.0$, $SD = 21.9$, to $M = 63.2$, $SD = 20.9$, respectively), $t(69) = 6.65$, $p < .001$.

### Effects of the Condition Manipulation on Intervention Activities

To better understand how condition impacted knowledge gains, we explored the effects of the condition manipulation on interven-

tion activities. Before reporting these effects, it is important to note that the manipulation did not impact the amount of material covered during the intervention; on average, students in the compare and sequential conditions studied approximately 20 of the 24 available worked examples ($M = 20.9$, $SD = 0.6$, vs. $M = 20.3$, $SD = 0.41$, respectively) and solved 10 of the available 12 practice problems ($M = 10.1$, $SD = 0.4$, vs. $M = 9.5$, $SD = 0.3$, respectively). In addition, students in both conditions were usually able to comprehend the individual solution steps; they correctly labeled almost all of the unlabeled steps in the worked examples (compare: $M = 97\%$, $SD = 5\%$, vs. sequential: $M = 92\%$, $SD = 15\%$). We expected the compare condition to support more explicit comparisons between multiple methods, including their accuracy and
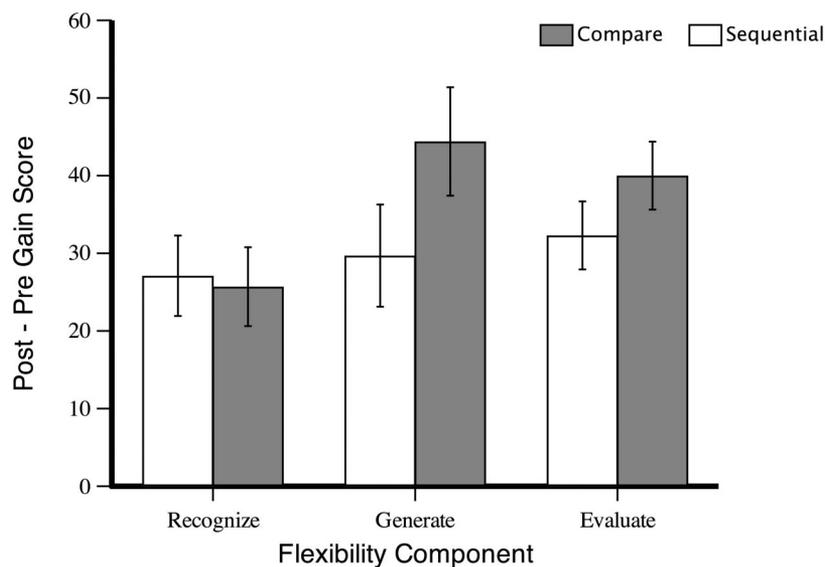


*Figure 3.* Flexibility gain for three components of flexibility, by condition (error bars are SE).

efficiency, and to support greater adoption of the shortcut strategies. Because of the exploratory nature of these analyses that required the use of multiple tests, we adopted the more conservative alpha value of .005 when interpreting the findings.

*Flexibility: Solution methods.*   On the practice problems, students in the compare condition were much more likely to adopt the demonstrated shortcut ($M = 41\%$ of intervention problems solved, $SD = 32\%$) than students in the sequential condition ($M = 13\%$, $SD = 28\%$). Multilevel modeling was used to confirm the effect of condition on use of shortcut methods. To better understand the role of prior knowledge in adopting the shortcut methods, we included procedural knowledge pretest scores for both the individual and his or her partner at the first level of the model. Students in the compare condition used the shortcuts much more frequently during the intervention, $t(31.0) = 4.70$, $p < .001$, $d = .93$. Their partners' procedural knowledge at pretest also positively impacted use of shortcuts during the intervention, $t(63.2) = 3.65$, $p < .001$, but their own pretest score did not.

*Explanation quality.*   Student pairs also provided written explanations to reflection questions when studying the worked examples. Students in the two conditions answered different questions, and our two coding schemes were designed to indicate whether our condition manipulation had its intended effects. The first coding scheme focused on four general characteristics of the explanations, as shown in Table 6. Children in the compare condition almost always referenced multiple methods, focused on the solution method, and judged the efficiency or accuracy of the methods. They sometimes mentioned the shortcut step or used mathematical terms to justify their ideas. A representative explanation from a pair in the compare condition was: "It is OK to do either step if you know how to do it. Mary's way is faster, but only easier if you know how to properly combine the terms. Jessica's solution takes longer, but is also OK to do." In contrast, children in the sequential condition referenced multiple methods much less often, focused less on the method, and were less likely to judge the efficiency of solutions. A representative response from a pair in the sequential group was: "Yes [I would choose this way] because he distributed in the right way and he added and divided on both sides correctly." Typically, students in the sequential condition focused on justifying a single solution method.

Looking more closely at the written explanations that referenced multiple solutions, the second coding scheme focused on explicit use of comparison along three dimensions (see Table 7). As expected, students in the compare condition were more likely to make explicit comparisons. In particular, students in the compare group were more likely to compare answers and/or to note the difference in the efficiency of the steps in the two solution methods. Students in the sequential condition rarely did these things, as they were difficult to do without side-by-side comparison of multiple methods to the same problem. Surprisingly, students in both conditions were equally likely to compare methods, either by directly comparing solution steps (typically in the compare condition) or by suggesting alternative methods (typically in the sequential condition).

Overall, students' explanations confirmed that the condition manipulation had its intended effect. Students in both conditions generated coherent explanations. The format and questions used in the compare condition elicited consideration of multiple methods and comparative judgments of the accuracy and efficiency of the methods. The format and questions used in the sequential condition focused attention on a single method or answer and elicited less judgment and greater use of mathematical terminology.

We explored whether individual differences in the frequency of making explicit comparisons during the intervention predicted outcomes at posttest. In this model, frequency of generating comparisons during the intervention, rather than condition, was used as a predictor. Making more comparisons during the intervention was marginally predictive of procedural knowledge gain, $t(30.0) = 2.53$, $p = .017$. However, it did not reliably predict flexibility gain or conceptual knowledge gain ($ps = .18$).

*Partner interaction.*   The discussions of a pair of high-learning students and a pair of modest-learning students were transcribed to better understand how comparison supported learning. Their discussion of a set of worked examples is presented in Table 8. As these examples indicate, both the high-learning and modest-learning pairs carefully studied the worked examples and had a high level of turn-taking and engagement. Throughout the sessions, the high-learning pair noticed each key feature of the problems; worked to make sense of shortcut steps; compared the solution steps; and evaluated the accuracy, efficiency, and constraints of the methods. In contrast, the modest-learning pair did not generate comparisons, rejected all nonstandard solutions as inaccurate, and did not consider efficiency. These interactions illustrate the benefits of comparing multiple solution methods for

Table 6

*Percentage of Intervention Explanations Containing Each Feature, by Condition*

| Explanation characteristic | Sample explanation | Compare (%) | Sequential (%) | Cohen's *d* |
|---|---|---|---|---|
| Reference multiple methods | "It is okay to do it either way." | 92 | 25*** | 5.49 |
| Focus | | | | |
|   On method | "He divided each side by 2." | 90 | 77*** | 1.87 |
|     Shortcut | "Mary combined like terms." | 11 | 4* | 0.82 |
|   On answer | "The answer is right." | 29 | 27 | 0.16 |
| Judge | | | | |
|   Efficiency | "Jame's way was just faster." | 47 | 37* | 0.83 |
|   Accuracy | "Sammy's solution is also correct because she distributed correctly." | 32 | 26 | 0.50 |
| Justify mathematically | "Used the right properties at the right times." | 30 | 46* | −0.78 |

*Note.*   Differences between conditions were significant, with $df = 1, 31$, as marked (and adopting more conservative alpha values due to multiple tests).
* $p < .05$.   *** $p < .001$.

Table 7

*Percentage of Intervention Explanations Containing Comparisons, by Condition*

| Explanation characteristic | Sample explanations | Compare | Sequential | Cohen's $d$ |
|---|---|---|---|---|
| Compare methods | "Jessica distributed and Mary combined like terms." or "You could have combined first." | 11 | 12 | −0.10 |
| Compare answers | "They end up with the same answer after all the steps." | 16 | 0*** | 1.56 |
| Compare efficiency of steps | "Jill used more steps." | 19 | 2*** | 2.11 |
| Any comparison | At least one of the above done | 41 | 12*** | 2.23 |

*Note.* Differences between conditions were significant, with $df = 1, 31$, as marked.
*** $p < .001$.

pushing acceptance of nonstandard methods and highlighting differences in efficiency.

## Discussion

Comparing and contrasting alternative solution methods led to greater gains in procedural knowledge and flexibility, and comparable gains in conceptual knowledge, compared to studying multiple methods sequentially. The compare condition facilitated attention to and adoption of nonconventional methods by guiding attention to solution accuracy and efficiency. These findings provide direct empirical support for one common component of reform mathematics teaching. The present study also suggests that prior cognitive science research on comparison as a basic learning mechanism (e.g., Gentner et al., 2003; Namy & Gentner, 2002; Schwartz & Bransford, 1998) may be generalizable to a new domain (algebra), a new age group (school-age children), and a new setting (the classroom).

These findings were strengthened by the use of a unique methodology. At present there is a push by the federal government to

Table 8

*Sample Dialogue of a High-Learning and a Modest-Learning Pair During the Intervention*

| High learners: Ben and Krista | Modest learners: Allison and Matt |
|---|---|
| [Quickly describe Mandy's solution and move to Erica's solution]<br>Krista: "What'd they [Erica] do?"<br>Ben: "Subtracted $3(y + 1)$ and they had that as one whole term, so they…and then over here was $(y + 1)$. Subtracted $3(y + 1)$ from $5(y + 1)$ to get $2(y + 1)$. And this wasn't over here, so $2(y + 1) = 8$."<br>Ben: "That's correct. Subtracted them on both. So then $y + 1 = 4$, they divided this by two and divided this by two…. These are both correct."<br>Krista: "I believe, because when they divided it by two, what happened to, they just divided it by two and that kinda makes the two go bye-bye? Or"<br>Ben: "Because if you have two of this and you divide by two, you only have one $y + 1$, correct? And over here you divide 8 by two and have four.<br>Krista: "Right. Or you could also multiply by the reciprocal and basically get the same thing."<br>[They read the first question and clarify its meaning.]<br>Krista: "They both did the problem correctly…. But they just did different ways, but they got the same answer…. Mandy just kinda did a few extra steps, I believe. She did like"<br>Ben: "Mandy distributed"<br>Krista: "and combined."<br>Ben: "but over here, Erica used, she like"<br>Krista: "just went right on to subtraction."<br>Ben: "she used $3(y + 1)$ as a term for - I don't know, how would you say that?"<br>[They paused and asked the teacher for help. He helped them remember the phrase "like terms."]<br>Ben: "Mandy distributed and"<br>Krista: "combined…"<br>Ben: "and Erica subtracted. Subtracted"<br>Krista: "subtracted from the like terms. But then, they basically did the same steps after that, but just in a different order."<br>[Finally, prompted by the second question, after a brief discussion,]<br>Krista concluded: "It's quicker…more efficient." | [Explaining Mandy's solution in Figure 1]<br>Allison: "So $5y+$, ok so she distributed $5y$, 5" Krista: "Oh, I getcha."<br>Matt: "Then she combined."<br>Allison: "No, yeah, no she distributed on both sides and combined"<br>Matt: "Yeah, she did."<br>Allison: "Where? Yeah, she combined on this side. You're right, she did."<br>Matt: "Exactly."<br>Allison: "She combined these, those two on that side. And then she subtracted on both. She subtracted $3y$ on both."<br>Matt: "And then she subtracted again and then she subtracted 5 on both sides."<br>Allison: "Oh, yeah. OK." [reads prompt] "Would you choose to use Mandy's way to solve the problem?"<br>Matt: "Yes."<br>Allison: "Because she used all the steps in the right way, and she combined. Yeah."<br>[Start next page]<br>Allison: "Erica. Here's one student's solution."<br>Matt: "Well, she did not distribute."<br>Allison: [Begins to read question] "Check Erica's solution…so let's pretend… $10x+30$ equals $6x+18$…she didn't get the right answer…"<br>Matt: "Yeah, so, no."<br>Allison: "No, she didn't distribute."<br>Matt: "She didn't distribute at all,"<br>Allison: "which gave her the wrong answer."<br>Matt: "OK."<br>Allison: "…and she didn't combine like terms." [They do not substitute answer into equations.] |

*Note.* This is a discussion of Erica's and Mandy's solutions shown in Figure 1.

use randomized field trials to demonstrate the efficacy of educational interventions (National Research Council, 2002). Such randomized trials are considered the gold standard, but this methodology is challenging to implement in classroom settings. Our use of random assignment of students to condition within their regular classroom context, along with maintenance of a fairly typical classroom environment, allowed us to provide experimental evidence on the efficacy of our approach. Further, rather than comparing our intervention to standard classroom practice, which differs from our intervention on many dimensions, we compared it to a control condition that was matched on as many dimensions as possible. This allowed us to evaluate a specific component of effective teaching and learning. We consider the implications of the current research for why comparison facilitates learning and for educational practice.

## Why Comparing Multiple Solutions Facilitates Learning

Comparing contrasting solutions seemed to support gains in procedural knowledge because it facilitated students' exploration and use of alternative solution methods. During the intervention, students in the compare condition were twice as likely to use the demonstrated shortcut to solve the practice problems. At posttest, they continued to be more likely to use the shortcut method and less likely to use the conventional method. In fact, greater use of shortcuts at posttest helped to explain the relation between condition and accuracy. Students in the compare condition also made greater gains on an independent measure of flexibility. For example, these students were better able to justify why a shortcut step was a good way to solve a particular equation. These findings suggest that comparison of multiple solutions helped students move beyond rigid adherence to a single solution method to more adaptive and flexible use of multiple methods.

How might comparing contrasting solutions support greater procedural knowledge and flexibility? First, it seems to help students differentiate important problem features (e.g., notice the shortcut step and the efficiency of methods; see Table 6; Schwartz & Bransford, 1998). Second, it seemed to help students consider multiple methods in general. Finally, it may better prepare students to learn from a summary lesson presented to all students (Schwartz & Bransford, 1998).

We expected that comparison students' greater improvements in procedural knowledge would be accompanied by greater improvements in conceptual knowledge (Baroody & Dowker, 2003; Rittle-Johnson et al., 2001). Students in the compare condition were better able to transfer their methods to novel problems, suggesting these students may have greater conceptual knowledge. However, students in both conditions made modest gains on our independent measure of conceptual knowledge, and there was no difference between conditions in amount of gain. In fact, analyses of students' explanations during the intervention indicated that the compare condition reduced mathematical justifications (arguably another indicator of conceptual knowledge). At the same time, frequency of generating comparisons during the intervention was not predictive of conceptual gain. Together, these findings do not support our hypothesis that comparison of multiple solution methods would lead to improved conceptual knowledge. However, we suspect that four revisions to our method would lead comparison to support greater conceptual knowledge: (a) revising our reflection prompts to focus more on the concepts justifying different solutions methods; (b) including teacher-led discussion that highlights underlying concepts; (c) increasing the time of the intervention and covering a wider variety of problems; and (d) revising our conceptual knowledge measure to more directly assess concepts highlighted by comparison, such as composite variables.

## Implications for Reform Efforts in Mathematics Education

The current study provides the first experimental evidence supporting the benefit of actively encouraging comparison in mathematics education, at least for private middle-school students learning to solve equations. Students who studied and reflected on varied solution methods one at a time did not learn as much as those who compared and contrasted two methods to the same problem. Simple exposure to multiple ways may not maximize learning, underscoring concerns that some teachers' attempts to implement reform pedagogy have resulted in simple show-and-tell of student methods without discussion or comparison of the methods (Ball, 2001; Chazan & Ball, 1999).

How can comparison be supported in the classroom? Past research suggests that three features of our intervention materials may have been particularly important. First, a written record of all to-be-compared solution methods may be needed, preferably with the solution steps aligned (Fraivillig et al., 1999; Richland, Zur, & Holyoak, 2007). Second, explicit opportunities to identify similarities and differences in methods seems critical; it is encouraged by expert mathematics teachers (Fraivillig et al., 1999; Huffred-Ackles et al., 2004; Lampert, 1990; Silver et al., 2005) and leads to improved transfer in laboratory tasks (Catrambone & Holyoak, 1989; Gentner et al., 2003). Finally, instructional prompts may be needed to encourage students to consider the efficiency of the methods (Fraivillig et al., 1999; Lampert, 1990).

In the current study, scaffolds for effective comparison were embedded in the instructional materials, rather than being provided verbally by the teacher. Indeed, carefully crafted explanation prompts, worked examples, and peer collaboration seemed to support productive explanation during partner work in the classroom (e.g., Fuchs et al., 1997); contrasting examples with explicit comparison prompts may be one way to support effective explanation. Nevertheless, we suspect that teacher-led, whole-class discussion would further enhance these benefits.

## Limitations and Future Directions

The current study was an important first step in providing experimental evidence for the benefits of comparing alternative solution methods, but much is yet to be done. First, it is critical to replicate these findings under more typical conditions in public school classrooms, including students with more diverse abilities and teachers with diverse teaching styles, fewer classroom resources, and larger class sizes. For example, comparison of alternative solution methods is likely only effective for students with sufficient prior knowledge (e.g., conceptual and procedural knowledge for solving one- and two-step equations). In addition, in the present study the teacher may have augmented the effects of our manipulation via verbal prompts and explanations to students during partner work.

Next, it is important to evaluate when and how comparison facilitates learning. Is comparison effective across a variety of mathematical topics and with a wider range of ages and mathematical abilities? Are some types of comparisons more important than others? How can teachers extend student-generated comparisons during whole-class discussions? Does comparison lead to lasting change on standardized measures? We are in the process of addressing several of these issues in new studies. First, we are working in pre-algebra classes in public schools and examining the effectiveness of different types of comparisons. Second, we are working with younger students (fifth-graders) in a very different mathematical domain (computational estimation). In both of these studies, we hope to confirm the present findings and to better evaluate the impact of comparison on conceptual knowledge gain.

## Conclusion

Comparison seems to be a fundamental learning process. In particular, comparing multiple methods to the same problem facilitates learning, particularly procedural knowledge and flexibility. Moving beyond simple show-and-tell of different solution methods to more active sharing-and-comparing is an important goal in reform efforts in mathematics. This study provides direct empirical evidence that in learning to solve equations, it pays to compare.

## References

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70,* 181–214.

Ball, D. L. (1993). With an eye on the mathematical horizon: Dilemmas of teaching elementary school mathematics. *The Elementary School Journal, 93,* 373–397.

Ball, D. L. (2001). Teaching, with respect to mathematics and students. In T. Wood, B. Scott Nelson, & J. Warfield (Eds.), *Beyond classical pedagogy: Teaching elementary school mathematics* (pp. 11–22). Mahwah, NJ: Erlbaum.

Ballheim, C. (1999, October). Readers respond to what's basic. *Mathematics Education Dialogues, 3,* 11.

Baroody, A. J., & Dowker, A. (2003). *The development of arithmetic concepts and skills: Constructing adaptive expertise.* Mahwah, NJ: Erlbaum.

Barzi, F., & Woodward, M. (2004). Imputations of missing values in practice: Results from imputations of serum cholesterol in 28 cohort studies. *American Journal of Epidemiology, 160,* 34–45.

Beishuizen, M., van Putten, C. M., & van Mulken, F. (1997). Mental arithmetic and strategy use with indirect number problems up to one hundred. *Learning and Instruction, 7,* 87–106.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13,* 221–252.

Blöte, A. W., Van der Burg, E., & Klein, A. S. (2001). Students' flexibility in solving two-digit addition and subtraction problems: Instruction effects. *Journal of Educational Psychology, 93,* 627–638.

Blume, G. W., & Heckman, D. S. (1997). What do students know about algebra and functions? In P. A. Kenney & E. A. Silver (Eds.), *Results from the sixth mathematics assessment* (pp. 225–277). Reston, VA: National Council of Teachers of Mathematics.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1147–1156.

Chazan, D., & Ball, D. (1999). Beyond being told not to tell. *For the Learning of Mathematics, 19,* 2–10.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439–477.

Dowker, A. (1992). Computational estimation strategies of professional mathematicians. *Journal for Research in Mathematics Education, 23,* 45–55.

Educational Records Bureau. (2004). *Comprehensive Testing Program, 4th edition.* New York: Educational Testing Service.

Fraivillig, J. L., Murphy, L. A., & Fuson, K. (1999). Advancing children's mathematical thinking in everyday mathematics classrooms. *Journal for Research in Mathematics Education, 30,* 148–170.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., Karns, K., & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal, 97,* 223–249.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95,* 393–405.

Grawitch, M. J., & Munz, D. C. (2004). Are your data nonindependent? A practical guide to evaluating nonindependence and within-group agreement. *Understanding Statistics, 3,* 231–257.

Harrell, F. E. (2001). *Regression modeling strategies with applications to linear models, logistic regression, and survival analysis.* New York: Springer.

Hiebert, J. (1986). *Conceptual and procedural knowledge: The case of mathematics.* Hillsdale, NJ: Erlbaum.

Hiebert, J., & Carpenter, T. (1992). Learning and teaching with understanding. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). New York: Simon & Schuster Macmillan.

Huffred-Ackles, K., Fuson, K., & Sherin Gamoran, M. (2004). Describing levels and components of a math-talk learning community. *Journal for Research in Mathematics Education, 35,* 81–116.

Johnson, D. W., & Johnson, R. T. (1994). *Learning together and alone: Cooperative, competitive and individualistic learning* (4th ed.). Boston: Allyn and Bacon.

Kenny, D. A., Kashy, D. A., Mannetti, L., Pierro, A., & Livi, S. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology, 83,* 126–137.

Kieran, C. (1992). The learning and teaching of school algebra. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390–419). New York: Simon & Schuster.

Kilpatrick, J., Swafford, J. O., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Kurtz, K., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *The Journal of the Learning Sciences, 10,* 417–446.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27,* 29–63.

Larson, R., Boswell, L., Kanold, T., & Stiff, L. (1999). *Passport to algebra and geometry.* Evanston, IL: McDougal Littell.

Loewenstein, J., & Gentner, D. (2001). Spatial mapping in preschoolers: Close comparisons facilitate far mappings. *Journal of Cognition and Development, 2,* 189–219.

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General, 131,* 5–15.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics.* Reston, VA: Author.

National Research Council. (2002). *Scientific research in education.* Washington, DC: National Academy Press.

Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy, 7,* 85–98.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research, 74,* 525–556.

Richland, L. E., Zur, O., & Holyoak, K. J. (2007, May 25). Cognitive supports for analogies in the mathematics classroom. *Science, 316,* 1128–1129.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93,* 346–362.

Satterthwaite, F. E. (1946). An approximate distribution of estimation of variance components. *Biometrics Bulletin, 2,* 110–114.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Schmidt, W. H., McKnight, C. C., Cogan, L. S., Jakwerth, P. M., & Houang, R. T. (1999). *Facing the consequences: Using TIMSS for a closer look at US mathematics and science education.* Dordrecht, the Netherlands: Kluwer.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16,* 475–522.

Silver, E. A., Ghousseini, H., Gosen, D., Charalambous, C., & Strawhun, B. (2005). Moving from rhetoric to praxis: Issues faced by teachers in having students consider multiple solutions for problems in the mathematics classroom. *Journal of Mathematical Behavior, 24,* 287–301.

Star, J. R. (2005). Reconceptualizing procedural knowledge. *Journal for Research in Mathematics Education, 36,* 404–411.

Star, J. R. (2007). Foregrounding procedural knowledge. *Journal for Research in Mathematics Education, 38,* 132–135.

Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology, 31,* 280–300.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom.* New York: Free Press.

Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education* (6th ed.). Columbus, OH: Merrill Prentice-Hall.

Webb, N. M. (1991). Task-related verbal interaction and mathematics learning in small groups. *Journal for Research in Mathematics Education, 22,* 366–389.

---

## E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at http://watson.apa.org/notify/ and you will be notified by e-mail when issues of interest to you become available!