Compared to what? The effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving

Share Your Story

Compared to what?

The effects of different comparisons on conceptual knowledge

and procedural flexibility for equation solving

Bethany Rittle-Johnson

Vanderbilt University

Jon R. Star

Harvard University

Bethany Rittle-Johnson, Department of Psychology and Human Development, Peabody College, Vanderbilt University. Jon R. Star, Graduate School of Education, Harvard University.

Correspondence concerning this article should be addressed to Bethany Rittle-Johnson, 230 Appleton Place, Peabody #0552, Nashville, TN 37203, 615-322-8301 (voice), b.rittle-johnson@vanderbilt.edu

Abstract

Researchers in both cognitive science and mathematics education emphasize the importance of comparison for learning and transfer. However, surprisingly little is known about the advantages and disadvantages of what types of things are being compared. In this experimental study, 162 7th- and 8th-grade students learned to solve equations by comparing equivalent problems solved with the same solution method, by comparing different problem types solved with the same solution method, or by comparing different solution methods to the same problem. Students' conceptual knowledge and procedural flexibility were best supported by comparing solution methods, and to a lesser extent by comparing problem types. The benefits of comparison are augmented when examples differ on relevant features, and contrasting methods may be particularly useful in mathematics learning.

KEYWORDS: Learning processes; Conceptual (declarative) knowledge; Procedural knowledge; Mathematics education; Mathematics concepts

Compared to what? The effects of different comparisons on conceptual knowledge and

procedural flexibility for equation solving

Experts agree; comparison is good. Researchers in both cognitive science and

mathematics education emphasize the importance of comparison for learning and transfer (e.g.

Ball, 1993; Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1983; Silver,

Ghousseini, Gosen, Charalambous, & Strawhun, 2005). As Gentner (2005) recently noted, "The

simple, ubiquitous act of comparing two things is often highly informative to human learners….

Comparison is a general learning process that can promote deep relational learning and the

development of theory-level explanations" (pp. 247, 251). Despite widespread agreement on the

merits of comparison, surprisingly little is known about the advantages and disadvantages of

what types of things are being compared. Designing effective educational interventions requires

a better understanding of what should be compared. In turn, evaluating alternative interventions

reveals new components of comparison that need to be incorporated into theories of learning.

In the introduction, we briefly overview experimental research on comparison and

descriptive research on the use of comparison in mathematics classrooms. Next, we identify

limitations in prior research on comparison for designing educational interventions and identify

three potential types of comparisons that could be used. Finally, we overview our target domain

and outcomes and how we evaluated the three types of comparison for supporting middle-school

students' conceptual and procedural knowledge of equation solving.

Experimental Research on Comparison

Experimental studies on comparison have yielded three key findings: 1) two examples are

better than one (Gick & Holyoak, 1983; Namy & Gentner, 2002), 2) two examples presented

together are better than two examples presented separately (Gentner et al., 2003; Oakes & Ribar, 2005), and 3) instructional support augments the benefits of comparison (Catrambone & Holyoak, 1989; Gentner et al., 2003; Schwartz & Bransford, 1998; Tennyson & Tennyson, 1975; VanderStoep & Seifert, 1993). Consider the classic example on analogical reasoning using the Dunker Radiation problem. After studying an isomorphic problem and its solution, participants were told they were doing a new experiment and were given the Dunker Radiation problem to solve. Without explicit hints, they were very unlikely to recognize that the solution given in the isomorphic problem could be used to solve the radiation problem (Gick & Holyoak, 1980). However, providing a solution to two isomorphic problems with different surface features and incorporating instructional support by prompting for comparison greatly increased spontaneous generalization of the solution to the Dunker Radiation problem (Catrambone & Holyoak, 1989; Gick & Holyoak, 1983).

Largely absent from this extensive cognitive science literature are investigations of the benefits of comparison in school-aged children or with academic tasks. Recently, our work has attempted to address both of these limitations. Seventh graders compared pairs of worked-examples for algebra equations or studied the same worked-examples one at a time during partner work in their mathematics classes. Students in the comparison condition had greater procedural knowledge and flexibility (Rittle-Johnson & Star, 2007). We have replicated these findings for fifth graders learning about computational estimation (Star & Rittle-Johnson, 2008).

Comparison in Mathematics Classrooms

Despite the relative lack of experimental research on comparison with school-age children or with academic tasks, mathematics educators are aware of the benefits of comparison and have attempted to incorporate its use in instruction. Comparing, reflecting on, and discussing

multiple solution methods is thought to improve student learning (Silver et al., 2005). Expert

mathematics teachers have students share and compare solution methods (e.g. Ball, 1993;

Lampert, 1990) and this instructional practice is emphasized in the National Council of Teachers

of Mathematics (NCTM) *Standards* (1989; 2000). Furthermore, comparison is not just used by

expert teachers. Analyses of the 8th-grade videos from the Trends in International Mathematics

and Science Study (TIMSS) found that a representative sample of teachers in the US, Japan and

Hong Kong all made comparisons multiple times in their lessons (Richland, Holyoak, & Stigler,

2004; Richland, Zur, & Holyoak, 2007).

However, US teachers have been challenged in their attempts to use comparison

effectively. Analyses of the TIMSS videos indicated that, while US teachers were making

comparisons, they were frequently not made in ways that seem most conducive to student

learning (Richland et al., 2004; Richland et al., 2007). For example, teachers, rather than

students, were usually initiating the comparisons and making the links between examples

(Richland et al., 2004). US teachers also made comparisons to non-mathematical contexts with

little instructional support. In laboratory studies, people rarely apply information learned in one

context to another context if the two contexts do not share surface similarities (Gick & Holyoak,

1983; Reed, Ackinclose, & Voss, 1990), raising concerns that students are not learning from

these comparisons. Other studies have shown that many teachers merely present (or have

students present) multiple examples, without any instructional support or linkages between

examples (Chazan & Ball, 1999). Simply suggesting that teachers use comparison in

mathematics teaching does not appear to have been successful in promoting US teachers'

effective use of this instructional approach.

There is a clear need for research that helps bridge between experimental research on comparison and successful implementation of comparison in the classroom; our prior work is one such effort (Rittle-Johnson & Star, 2007; Star & Rittle-Johnson, 2008). However, in our attempts to help teachers use comparison to improve students' learning of mathematics, it also became clear that there are significant gaps in the cognitive science literature on comparison – gaps that are critical to teachers' successful implementation of comparison.

Taking Cognitive Science to the Classroom: Current Limitations

Research from cognitive science on comparison provides some guidelines for how to effectively use comparison in the classroom, such as the merits of comparing worked examples and the benefit of prompts that guide attention to important comparisons (Catrambone & Holyoak, 1989; Gentner et al., 2003). However, attempting to translate and incorporate cognitive science research on comparison into instruction reveals limitations in the existing literature that need to be addressed. In particular, the cognitive science literature provides limited guidance on one of the most important decisions that must be made in the implementation of comparison – namely, *what* should be compared. When two examples are to be compared, what dimensions of the examples should vary and what dimensions should remain the same?

A global issue is how similar the two examples should be, particularly the surface features of the problems (e.g. the cover stories or problem format). Without high surface similarity in the problems, people often fail to notice the underlying similarities in problems, and thus do not spontaneously use a demonstrated solution to solve related problems (e.g. Gick & Holyoak, 1980; 1983 research on Dunker Radiation problem; also see Ross & Kennedy, 1990). However, there is a cost to high surface similarity; people may learn the solution but not generalize it to new problems with different surface features (Reed, 1989). Research using

categorization tasks indicates that comparing examples with moderate surface similarity, rather than high surface similarity, supports better generalization of the category (Tennyson, Tennyson, & Rothen, 1980; Tennyson & Tennyson, 1975; Waxman & Klibanoff, 2000). For problem-solving tasks, there is more limited evidence that moderately similar examples are preferable to highly similar examples for supporting abstraction or generalization of the problem-solving solution (Gick & Paterson, 1992; VanderStoep & Seifert, 1993). Similarity in examples seems to be an important factor, but it is difficult to specify how similar examples should be.

Furthermore, the existing literature is not clear on which dimensions of examples should be similar. This issue is of particular importance in mathematics education, where it is critical to consider the similarity of the methods used to solve the problems as well as the similarity of the problems themselves. For example, both the problems and the solution methods could be very similar or either the problems or the solution methods could differ. Prior work in cognitive science does not speak to the affordances and constraints of these different types of comparison because most work only involved comparison of similar problems with the same underlying solution method (i.e., isomorphic problems). Prior work in mathematics education (as well as our own research) has focused on the opposite type of comparison - comparing the same problem solved with two different methods. Ultimately, either problems or solution methods can vary in pairs of mathematics examples. The existing literature does not provide guidance on the benefits and drawbacks to varying these different dimensions.

The goal of the current study was to extend existing cognitive science research on comparison by evaluating three types of comparison for supporting mathematics learning. The three types varied in how the problems and solution methods differed (see Table 1). Two comparison types were based on what is used in cognitive science research, where learners

typically compare isomorphic problems with the same solution. For one type, the isomorphic

problems were the same problem type and thus had very similar surface features and solutions.

For the other type, the isomorphic problems had different problem features, and thus had

moderately similar surface features but the same underlying solution method. The third type,

which emerged from the mathematics education literature, was comparing different solution

methods to the same problem.

This work has several theoretical implications. First, it tests an implicit assumption in

cognitive science on the primary role of comparison in problem solving – that comparison

facilitates extraction of a common solution method from examples with different surface

features. This focus on comparing isomorphic examples grew out of research on analogical

problem solving (e.g. Gick & Holyoak, 1980) and has shifted attention away from the possibility

that comparing multiple solution methods to the same problem would also benefit learners.

Indeed, children and adults typically use multiple methods to solve problems and choosing

flexibly among methods accounts for major advances in problem-solving performance across a

variety of domains (Siegler, 1996). Thus, we predicted that comparing solutions methods would

be at least as effective, if not more effective, than comparing isomorphic problems, given the

importance of flexible choice among multiple methods and the ability of comparing solution

methods to support flexibility (Rittle-Johnson & Star, 2007). Second, our findings will reveal

potential constraints that can guide expansion of theories of analogical learning to include a

wider range of comparisons. In addition to these theoretical contributions, this research will help

shape practical guidelines that mathematics teachers can use to improve their use of comparison

in the classroom.

Target Domain and Outcomes

We evaluated the effectiveness of different types of comparison for learning a core

component of mathematics –linear equation solving. It is considered a "basic skill" by many in

mathematics education and is recommended as a Curriculum Focal Point for Grade 7 by NCTM

(Ballheim, 1999; National Mathematics Advisory Panel, 2008; NCTM, 2006). When introduced,

the methods used to solve equations are among the longest and most complex to which students

have been exposed. Thus, in equation solving, students need to learn multiple rules and heuristics

for how to combine the rules, rather than a single principle or rule as is typical in laboratory

studies (VanLehn, 1996).

We used three types of multi-step linear equations, taken from Rittle-Johnson & Star

(2007) (e.g. $3(x + 1) = 15$, see Table 2). In the center column of Table 2 is a conventional and

commonly taught method for solving linear equations that applies to most equations: distribute,

combine like terms, subtract constants and variables from both sides, and then divide both sides

by the coefficient. In the right-most column of the table is a non-conventional method that treats

expressions such as $(x + 1)$ as a composite variable and is arguably a shortcut – it is more

efficient because it involves fewer steps and fewer computations; thus it may be executed faster

and with fewer errors. For example, to solve $3(x + 1) = 15$, rather than distributing the 3, you can

divide both sides by 3. This non-conventional method can push children to understand important

problem features and reflect on when different methods are more efficient.

Our target outcomes were three critical components of mathematical competence:

procedural knowledge, procedural flexibility, and conceptual knowledge (Hiebert, 1986;

Kilpatrick, Swafford, & Findell, 2001). First, procedural knowledge is the ability to execute

action sequences to solve problems (Hiebert, 1986; Rittle-Johnson, Siegler, & Alibali, 2001).

The adaptability of solution methods can be tested by novel transfer problems; problems that can be solved by modifying learnt solution methods to new problem features (Paas & Van Merrienboer, 1994; Singley & Anderson, 1989). Therefore, our procedural knowledge measure included both familiar and novel problem types. Second, procedural flexibility incorporates knowledge of multiple ways to solve problems and when to use them (Kilpatrick et al., 2001; Star, 2005) and is an important component of mathematical competence (Beishuizen, van Putten, & van Mulken, 1997; Blöte, Van der Burg, & Klein, 2001; Star & Seifert, 2006). To disentangle *knowledge* from *use*, we included an independent measure of flexibility knowledge as well as coded for flexible use of solution methods on the procedural knowledge assessment (Star & Rittle-Johnson, 2007). Finally, conceptual knowledge is "an integrated and functional grasp of mathematical ideas" (Kilpatrick et al., 2001, p. 118). We measured conceptual knowledge via the ability to recognize and to explain key concepts in the domain, in line with past measures of conceptual knowledge (e.g.Carpenter, Franke, Jacobs, Fennema, & Empson, 1998; Hiebert & Wearne, 1996; Rittle-Johnson & Alibali, 1999; Rittle-Johnson & Star, 2007).

Current Study: Effects of Different Types of Comparison on Student Learning

We evaluated three types of comparison for supporting middle-school students' learning about multi-step linear equations. Students in all conditions compared worked examples and answered questions about the examples with a partner during three classroom lessons. They studied an equal number of worked examples for each problem type and each solution method across the conditions. The three conditions differed in how the problems and solution methods varied *within* an example pair (see Table 1 and Figure 1). Students were randomly assigned to either: (1) compare equivalent problems solved with the same solution method (2) compare different problem types solved with the same solution method, or (3) compare the same problem

solved with two different solution methods. The first and second conditions were both forms of comparing problems with the same solution method, but they differed in how similar the problems were. In the compare equivalent condition, equations of the same type were paired (e.g. $2(x + 3) = 8$ and $5(y + 4) = 10$); in the compare problem types condition, different problem types were paired (e.g. $2(x + 3) = 8$ and $6(h + 1) = 3(h + 1) + 27$). In the latter case, students needed to look beyond obvious differences in surface features to identify the common solution method. In both conditions, half the example pairs illustrated the conventional solution method for both equations and the other half illustrated the shortcut method for both equations. In the third condition, the same problem was solved with the conventional method in one example and the shortcut method in the other example in the pair.

Based on the mathematics education literature and Rittle-Johnson and Star (2007), we hypothesized that comparing solution methods would lead to the greatest learning on all three outcomes. In this condition, students see the same solution steps being implemented in different ways. Consider the example in Figure 1 (panel A). Students see subtracting a term from both sides when the term is a composite variable (step 1 in Patrick's solution) and when it is a single variable (step 3 in Nathan's solution). The step labels should help students align these steps across examples and make two abstractions: 1) generalizing their concepts of variables and like terms (i.e. improved conceptual knowledge) and 2) generalizing the solution steps to include composite variables (i.e. improved procedural transfer). In addition, comparing the different sequencing of steps should help improve students' heuristics for combining steps, focusing their attention on the fact that there are multiple possible sequences and a particular sequence can improve efficiency (i.e. improved procedural flexibility).

However, the cognitive science literature suggests that either form of comparing

problems with the same solution method should also support learning. Comparing equivalent

equations should help students learn the basic solution steps, abstracting across the particular

numbers and letters used in a given equation (Catrambone & Holyoak, 1989; Gentner et al.,

2003; Gick & Holyoak, 1983). However, these solution steps may be linked to overly narrow

problem features (Reed, 1989). Thus, comparing equivalent equations may be less effective in

supporting procedural transfer, procedural flexibility or conceptual knowledge than the other two

conditions. Comparing problem types should lead to abstraction of fairly general solution steps

that should support some procedural transfer and flexibility (Gick & Paterson, 1992;

VanderStoep & Seifert, 1993).

<div align="center">Method</div>

Participants

Students were drawn from a rural public school, a suburban public school, and an urban

private school. All students in nine pre-algebra classes at the schools were invited to participate,

with a total of 162 students giving consent to participate (81 female). Students' mean age was

13.1 years (range: 11.9 years to 15.1 years) and a majority were Caucasian (5% African-

American, 5% Asian/Indian, and 1% Hispanic). Approximately 14% of students received free or

reduced lunches. The seventh (n = 114) and eighth graders were drawn from classes taught by

five different teachers. Students were tracked by ability for math class based on their

performance in math class the year before and their standardized test scores. Students were

drawn from 5 advanced (n = 91) and 4 regular mathematics classes. Within a school, students in

the regular and advanced classes used the same textbook, and in previous lessons, students had

learned about the distributive property, simplifying expressions, and solving one-step and simple

two-step equations.

Design

We used a pretest-intervention-posttest design, including a retention test. Pairs of

students within a classroom were randomly assigned to compare solution methods (abbreviated

as *methods*; $n = 54$), compare problem types (abbreviated as *problem types*; $n = 56$), or compare

equivalent equations (abbreviated as *equivalent*; $n = 52$). During the intervention, students

studied the worked-example pairs with a partner and answered explanation prompts designed to

guide attention to the example features targeted in each condition. Students also solved practice

problems and received mini-lectures during the intervention. The intervention occurred during

three consecutive mathematics classes.

Materials

Intervention. Packets of worked examples were created for each condition. On each page

of the packet, a pair of worked examples was presented side-by-side (see Figure 1). The packets

were as similar as possible. They all contained four instances of each of the three equation types

for a total of 12 worked examples. Half the worked examples illustrated the conventional

solution method and half illustrated the composite-variable shortcut method. The primary

difference between the packets was how the worked examples were paired. In the *methods*

packets, each worked-example pair contained the same equation, solved using the conventional

and shortcut method. These worked examples were primarily a subset of those used in Rittle-

Johnson & Star (2007). In the *problem types* packets, each worked-example pair contained two

different types of equations, each solved with the same method. For example, a combine

composite equation and a divide composite equation were shown together, each solved using the

shortcut method (see Figure 1). In the *equivalent* packets, each worked-example pair contained two instances of the same problem type solved with the same solution method. Across all packets, each solution step was labeled using one of four step labels (distribute, combine, add/subtract on both, multiply/divide on both). Students needed to complete the labels for most of the steps to encourage active processing of the examples. Past research indicates that common labels improve the benefits of side-by-side presentation of examples (Namy & Gentner, 2002).

Each pair of worked examples was presented along with two questions prompting students to compare and contrast the targeted dimensions for a given condition. Asking specific and detailed comparison questions leads to better learning than simple side-by-side presentation of examples with or without a generic prompt to compare them (Catrambone & Holyoak, 1989; Gentner et al., 2003). The questions were designed to tap five different levels of thinking, based on Bloom's taxonomy (comprehension, application, analysis, synthesis, and evaluation) (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956), and were equated as much as possible. As illustrated in Figure 1, questions in the *methods* condition focused on comparing the solution steps, including their feasibility and efficiency; those in both the *problem-types* and *equivalent* conditions focused on comparing both the problem features and the particular solution steps.

Each packet also included one guided practice problem, on which students were asked to use a particular shortcut method to solve a new equation, and four independent practice problems on which students could choose their solution methods. In the *methods* condition, students were asked to solve two practice problems each in two different ways, whereas four different equations were presented in the packets for the other conditions. In Rittle-Johnson & Star (2007), students did not solve guided practice problems and students in the methods condition were not prompted to solve the practice problems in two different ways; these adjustments were

made to increase use of the shortcut methods across conditions and to extend the methods

manipulation to the practice sets.

Three brief homework assignments were developed, primarily using problems in the

students' regular textbooks. The homework assignments had six problems each and were review

problems similar to those solved in class. They were the same across all conditions.

Assessment. The same assessment was used as an individual pretest, posttest, and

retention test. It was modified from the assessment used in Rittle-Johnson & Star (2007) and was

designed to assess conceptual knowledge, procedural knowledge, and procedural flexibility.

Sample items are included in Table 3. The nine conceptual knowledge items tapped students'

verbal and non-verbal knowledge of algebra concepts, such as equivalence, like terms, and

composite variables. Of the six items on the assessment in Rittle-Johnson and Star (2007), three

were dropped because they had low inter-item correlations, largely due to ceiling effects. Two

of the remaining three items were modified to focus more on understanding of composite

variables (e.g. the first sample item in Table 3). Three new items were added: an additional

equivalent expressions item involving composite variables and two items on like terms (see

Table 3).

The procedural knowledge measure assessed students' ability to solve equations, with

two mental math problems, three familiar problems, three near transfer problems, and two far

transfer problems. The mental math and familiar problems were the same types of problems as

those presented during the intervention, and thus could be solved using the same sequencing of

solution steps. The near transfer problems included a novel problem feature, such as additional

terms inside the parentheses, and could be solved by generalizing the solution steps or how they

were sequenced. The far transfer problems required using the same steps to transform the

equation to solve for a different variable, a task students had not done before. The mental and far

transfer problems had not been used in previous research; the familiar and near transfer problems

were similar to those used in Rittle-Johnson & Star (2007), but had been simplified to involve

easier calculations (e.g., using the coefficient 1/2 rather than -1/4).

Procedural flexibility was assessed in two ways. First, flexible *use* of solution methods

was assessed by whether students used efficient solution methods on the procedural knowledge

assessment (also called adaptive strategy choice (cf. Siegler, 1996)). Second, flexible *knowledge*

of solution methods was assessed on an independent measure. Flexibility knowledge items fell

into three categories: (a) ability to *generate* different solutions to an equation when prompted; (b)

ability to *recognize* appropriate first solution steps for a particular problem; and (c) ability to

evaluate innovative first solution steps for accuracy and efficiency. Unlike Rittle-Johnson and

Star (2007), the items involved near transfer problems, rather than familiar problem types, to

more rigorously assess procedural flexibility. We also expanded the number of generate

flexibility items.

Procedure

All data collection occurred within students' intact mathematics classes over five

consecutive classroom periods. The instruction replaced the students' regular instruction on

solving relatively complex linear equations (e.g., those involving distribution and variables on

both sides of the equation) and occurred immediately after regular instruction on solving basic

two-step linear equations. On Day 1, students completed the pretest. Students were given 40-50

minutes to complete the pretest, including 12 minutes to complete the first 8 procedural

knowledge items. Some time pressure was included for these items to encourage students to use

efficient solution methods.

On Day 2, an instructor gave a brief (10 minute) scripted introduction to students. The instructor was either one of the authors, a research assistant, or the regular classroom teacher, and all instructors followed a script. Instruction began with the class attempting to solve the equation $3(x + 1) = 12$ on their own. The instructor then worked through a solution together with the class using a conventional solution method. Class discussion focused on why the steps used in a given solution were OK to do. Then a model of appropriate work with a partner was demonstrated to show the students how to work through the packets.

Following this introduction, pairs of students began working on the packets. When studying the worked examples, students were instructed to describe each solution method to their partner and answer the accompanying questions first verbally, and then in writing. Each student had his or her own packet and wrote down answers after discussion with their partner. The written explanation served to push students to summarize their ideas and come to a consensus. On practice problems, students were asked to solve the problems on their own, compare answers with their partner, and have their answers checked by an adult. The classroom teacher and one or two members of the project team 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. All students were given the same homework assignment at the end of the class period.

Days 3 and 4 followed the same format, with a brief whole-class lesson introducing a new problem feature (variables on both sides on Day 3 and fractional coefficients on Day 4) followed by partner work on the packets for the day. Students started a new packet each day and did not return to finish incomplete packets from the previous day. At the end of Day 4, the

instructor provided a 8-minute wrap-up lesson that emphasized (1) there is more than one way to solve an equation, (2) any way is OK if the two sides of the equation are kept equal, and (3) some ways of solving equations are better or easier than others. Direct instruction augments the benefits of comparison (Schwartz & Bransford, 1998; Tennyson & Tennyson, 1975; VanderStoep & Seifert, 1993).

On Day 5, students were given 40-50 minutes to complete the posttest, which was identical in content and administration to the pretest. Two weeks later, all students completed the retention test, also identical to the pretest and posttest.

Coding

Assessment. On the procedural knowledge assessment, no one solved the far transfer problems correctly at pretest or posttest, so the items were dropped. The remaining 8 problems were scored for accuracy of the answer. In addition, students' solution methods were coded (except for the mental math items). For this coding, computational errors were ignored. We evaluated students' first solution step as (1) distributing across parentheses (the conventional solution method), (2) using a shortcut step that had been demonstrated in the worked examples (e.g., divide composite, combine composite, and subtract composite; see Table 2), (3) using an unusual or incorrect algebraic step, (4) using an informal, non-algebraic approach, such as guess-and-test or unwind, or (5) not attempting the problem. Frequency of using a shortcut method was used as an indicator of flexible use of procedures.

The flexibility knowledge assessment had three components. Percentage of possible points on each component was calculated, and the three percentages were averaged to yield an overall flexibility knowledge score. On the conceptual knowledge assessment, students received one point for each correct answer or explanation. See Table 3 for scoring details. Independent

coders coded the solution methods and explanation qualities across the assessment for 20% of

the sample, and exact agreement ranged from 86-89%. Discrepancies were discussed and codes

were altered when deemed appropriate by the primary coder.

     Intervention. We tallied how many practice problems each student completed during the

intervention (students found the correct solution before moving on, so accuracy was not scored).

We also coded whether students used the demonstrated shortcut step to solve each practice

problem, and inter-rater reliability on 20% of the sample was 95%.

     Student pairs also provided written explanations during the intervention. Two coding

schemes were developed to code these explanations, which 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 86% to 99%.

Data Analysis

     Some students were absent on an assessment day. Nine students did not complete the

pretest, two did not complete the posttest, and eight did not complete the retention test. One of

these students was absent for both the pretest and retention test and was dropped from the

analyses. For the remaining seventeen students with only one missing assessment, statisticians

strongly recommend the use of imputation, rather than the traditional approach of omitting

participants with missing data, because it leads to more precise and unbiased conclusions (Peugh

& Enders, 2004; Schafer & Graham, 2002). When the data is missing at random (confirmed by

Little's MCAR test: $\chi 2$ (312) = 44.297, $p > .99$) and no more than 5% of the data is missing,

simulation studies indicate that imputation leads to the same conclusions as when there is no

missing data (e.g. Barzi & Woodward, 2004). As recommended by Schafer and Graham (2002),

we used the expectation-maximization (EM) algorithm for Maximum Likelihood Estimation via

the missing value analysis module of SPSS. The students' missing scores were estimated from all non-missing values on continuous variables that were included in the analyses presented below.

Because children worked with a partner for the intervention, we calculated intraclass correlations to test for non-independence in partner scores on the posttest and retention test, controlling for the predictor variables (Kenny, Kashy, & Cook, 2006). For the most part, partners' scores were statistically independent, with partial intraclass correlations ranging from -.18 to +.01 (p's > .2). The one potential exception was partner scores on the conceptual knowledge measure at retention test (r (229)= -.18, p = .11). Because the data was largely independent, we report ANCOVA models given their greater familiarity to the reader. The findings were equivalent when we used multilevel modeling to account for nesting within dyads.

<div align="center">Results</div>

We first overview students' knowledge at pretest. Next, we report the effect of condition on students' knowledge at posttest and retention test. Finally, we examine how the manipulation impacted intervention activities, such as the characteristics of students' explanations.

Pretest Knowledge

Recall that our intervention occurred after students had completed classroom lessons on solving basic one- and two-step equations. Thus, at pretest, students had some algebra knowledge. As shown in Table 4, 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, students most often used a conventional solution method and left a fair number of the problems blank (see Table 5). Use of composite-variable shortcuts was rare and only 10% of students used a shortcut at least once at pretest. Procedural knowledge correlated with both

conceptual knowledge ($r(162) = 0.50$, $p < 0.001$) and flexibility knowledge ($r(162) = 0.41$, $p <$ .001), and flexibility and conceptual knowledge were also related ($r(162) = 0.62$, $p < .001$).

At pretest, there were no significant differences between conditions on the procedural or conceptual knowledge measures $F(2, 159) = 0.42$ and 0.32, respectively (see Table 4). Although students had been randomly assigned to condition, there was a marginal difference between conditions on the flexibility knowledge measure, $F(2, 159) = 2.86$, $p = .06$. As shown in Table 4, students in the *equivalent* condition scored a bit higher than the other two conditions. Males and females did not differ in success on the pretest measures.

Effect of Condition on Knowledge at Posttest and Retention Test

As shown in Figures 2 and 3, students in the methods condition had the greatest conceptual knowledge and procedural flexibility. Separate repeated-measures ANCOVAs were conducted for each outcome, with time of assessment as a within-subject factor (posttest and retention test) and condition as a between-subject factor. Pretest accuracy on each measure, school, and classroom ability group (regular or advanced math class) were included as covariates to control for prior knowledge differences. When there was a main effect for condition, least-significant difference tests were used to compare performance in the three conditions.

We expected performance to remain the same or improve from posttest to retention test because students continued to learn about equation solving in their classrooms after the conclusion of our intervention. However, we did not expect condition to interact with time, but rather to remain stable across posttest and retention test. We also did not expect condition to vary by ability group. We included a condition x ability group interaction term in the initial analyses, but the two did not interact, so the interaction term was not included in the final models.

Conceptual knowledge. Conceptual knowledge varied by condition (see Figure 2 and Table 6). As expected, students who compared solution methods had higher accuracy on conceptual knowledge items across the posttest and retention test than either of the other two groups (p's $\leq$ .01). Students who compared equivalent vs. different problem types did not differ ($p = .45$). Inspection of accuracy on each item suggested that comparing methods was most effective at supporting knowledge of composite variables, both identifying like terms and thinking about equivalent expressions involving composite variables.

Prior conceptual, procedural, and flexibility knowledge, as well as ability group, each positively predicted conceptual knowledge across the posttest and retention test. Finally, conceptual knowledge improved from posttest to retention test, but the effect of condition did not vary for the two time points.

Procedural knowledge. Procedural knowledge did not vary by condition (see Table 6). Rather, all groups showed improvements, and individual differences in prior knowledge and ability grouping were the main (positive) predictors of knowledge differences. Procedural knowledge did improve from posttest to retention test, but the effect of condition did not vary for the two time points. When considering accuracy on only the novel equations, there still were no effects for condition ($M = 45\%$, 41% and 51% correct for method, problem types and equivalent conditions, respectively), $F(2, 153) = 1.60$, $p = .21$. As expected, all three conditions were equally effective at supporting success on familiar problem types. Unexpectedly, the three conditions did not differ in their effectiveness at supporting transfer to novel problem features.

Flexible use of solution methods. Although students across the conditions had similar accuracy on the procedural knowledge assessment, they were not equally flexible in their use of solution methods. On the procedural knowledge items, the composite-variable shortcut was more

efficient than a conventional, distribute-first, method and thus indicated more adaptive and

flexible use of solution methods. Students who compared equivalent equations used the shortcut

steps less often than either of the other two groups (p's $\leq .01$), and students who compared

solution methods and problem types used the shortcut steps equally often ($p = .71$) (see Figure 3

and Table 6). In addition, prior conceptual knowledge, flexibility knowledge, and ability group

positively predicted frequency of shortcut use. Students also were more likely to use shortcut

steps at retention test than at posttest, but time did not interact with condition.

Students could learn three different shortcut steps (see Table 2). Most students used the

divide composite shortcut, and the percent of students who used it did not vary by condition

(70%, 66% and 65% of students used it at least once across the posttest and retention test for the

equivalent, problem types and methods conditions, respectively). A large number also used the

combine composite shortcut, and students in the methods and problem types condition were

somewhat more likely to use it than those in the equivalent condition (72%, 71% and 57% of

students, respectively), although the effect of condition did not reach significance, $p = .15$.

Fewer students used the subtract composite shortcut, and the number of students who used it

varied by condition, $\chi^2 (2) = 10.20$, $p = .006$. Students in the methods condition used it much

more than those in the equivalent condition (69% vs. 38% of students, respectively), $\chi^2 (1) =$

10.18, $p = .001$, and somewhat more than those in the problem types condition (52% of

students), $\chi^2 (1) = 3.201$, $p=.073$. Students in the problem types condition used in more than

those in the equivalent condition, but not significantly so, $p = .14$.

Using a shortcut method improved accuracy. Frequency of shortcut use was positively

related to problem-solving accuracy on the procedural knowledge assessments. Frequency of

using shortcuts at posttest was positively related to accuracy at posttest, $r(157) = .47$, $p < .001$,

and using shortcuts at retention test was positively related to retention accuracy, $r(157) = .39$, $p <$ .001, after controlling for pretest knowledge measures.

As expected, comparing equivalent equations was the least effective at supporting flexible use of solution methods. For the most part, comparing solution methods or problem types were equally effective at supporting flexible use. However, there was some indication that comparing solution methods was a bit more effective (i.e. more students used the divide composite shortcut and they used atypical algebra solution methods more often (see Table 5)).

Flexibility knowledge. In addition to evaluating flexible use of solution methods, previous research indicates that independent measures of flexibility knowledge are important, particularly because they are more sensitive to early emerging knowledge (Blöte et al., 2001; Star & Rittle-Johnson, 2007). Condition did impact students' flexibility knowledge (see Table 6). Students who compared solution methods scored higher than students who compared equivalent equations, $p < .002$, and marginally higher than students who compared problem types, $p < .069$. Students who compared problem types had greater flexibility knowledge than those who compared equivalent equations, but this difference did not reach significance ($p = .158$). Prior knowledge and ability group also had a strong, positive, influences on flexibility knowledge. Students' flexibility increased from posttest to retention test, but the effect of condition did not vary by assessment time.

Follow-up analyses on the three subscales of the flexibility knowledge measure indicated the same effects for condition on the generating multiple methods and evaluating non-conventional methods subscales. Although means were in the expected direction, effects of condition were not significant for the recognize multiple methods subscale.

In summary, comparing solution methods generally led to greater conceptual knowledge and procedural flexibility than comparing equivalent or different problem types. However, it did not lead to greater procedural knowledge. Comparing problem types was more effective than comparing equivalent problems for supporting flexibility, but not conceptual or procedural knowledge.

Effects of the Condition Manipulation on Intervention Activities

Students' explanations during the intervention served as a manipulation check and provided some insights into how the condition manipulation impacted knowledge change. It is worth noting that condition did *not* impact the amount of material covered during the intervention; on average, students in the three conditions studied all of the worked examples, answered all of the questions, and solved 11 of the 12 available practice problems. It also did not influence choice of solution methods during the intervention. Students chose to use composite-variable shortcut steps on 46% of the practice problems, and this did not vary by condition.

Students in the three conditions answered different explanation questions designed to facilitate the appropriate comparisons for each condition. Characteristics of students' explanations verified that each condition had its intended affect. Further, we used the frequency of different explanation qualities to verify that particular types of explanations were predictive of learning. 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.[1]

First, we coded for the type of comparisons students made. Students compared four features of the examples; efficiency of the methods, specific solution steps, problem features, or the answers, as described in Table 7. We successfully promoted comparison across the conditions; over 70% of students' explanations included explicit comparisons. However, the

frequency of different types of comparison varied by condition. Students in the methods

condition were the most likely to compare the efficiency of the methods. Students in the

problem-types condition were the most likely to compare solution steps, often noting that the

first step differed in the two examples. Those in the equivalent condition were the most likely to

compare problem features, most often focusing on surface features such as the difference in the

particular numbers or letters. Thus, as intended, the methods condition guided attention to the

relative efficiency of different solution methods and the equivalent condition guided attention to

variations in surface features. We did not anticipate that the problem-types condition would

guide more attention to comparing solution steps than to comparing problem features; the

explanation prompts were very similar for the problem types and equivalent conditions. Side-by-

side presentation of different problem types encouraged greater comparison of solution methods

than of problem features.

　　　　To gain a better understanding of children's thinking during the intervention, we also

coded four general characteristics of the explanations. We coded the focus of the explanations,

references to multiple methods, evaluations of the examples, and use of mathematical

terminology, as described in Table 8. Students in the methods condition usually focused on the

solution methods and referenced multiple methods. They were most likely to evaluate the

examples, particularly the efficiency of the methods, but rarely used mathematical terms to

justify their ideas. A representative explanation in the methods condition was: "Nathan's way is

longer. Patrick doesn't distribute but Nathan does." Students in the problem-types condition also

focused on the solution method and referenced multiple methods. They evaluated both the

efficiency and problem features in the examples, but were less likely to make evaluations than

students in the methods condition, particularly the efficiency of the methods. For example, one

student wrote "She combines first and he subtracts first." Explanations in the equivalent

condition were more distinct. Students in this condition divided their focus across the method

and problem features and referenced multiple methods less than half the time. They often did not

evaluate the examples, although they were the most likely to evaluate the problem features. They

were also *most likely* to justify their ideas using mathematical terminology. For example, a

student in the equivalent condition explained: "Both [problems] have variables on both sides of

the equation, but one uses *y* and one uses *x*."

To better understand why the *methods* condition led to the greatest learning, we evaluated

which of the predominant features of their explanations (e.g., comparing efficiency, comparing

solution steps, focusing on methods, referencing multiple solutions, and evaluating efficiency)

predicted performance at posttest and retention test. In the models, frequency of each of the

explanation types, rather than condition, was used as a predictor. Comparing efficiency more

often positively predicted conceptual knowledge, $F(1, 154) = 9.346$, $p = .003$, $\eta^2 = .06$, and to a

lesser extent, flexibility knowledge, $F(1, 154) = 5.871$, $p = .017$, $\eta^2 = .04$. Referencing multiple

solutions more often positively predicted flexibility knowledge, $F(1, 154) = 7.885$, $p = .006$, $\eta^2 =$

.05, use of short cuts, $F(1, 154) = 9.741$, $p = .002$, $\eta^2 = .06$, and to a lesser extent, conceptual

knowledge, $F(1, 154) = 6.258$, $p = .013$, $\eta^2 = .04$. The other three features were not

systematically related to the outcomes, nor were any other types of comparisons. These

exploratory findings suggest that consideration of multiple solutions may have helped support

procedural flexibility and that specific comparisons of efficiency were particularly useful for

supporting conceptual knowledge.

Discussion

All comparisons are not equally effective. In the case of equation solving, comparing solution methods was more effective for supporting conceptual knowledge and procedural flexibility than comparing equivalent equations or comparing problem types. The methods condition pushed students to consider multiple solution methods and compare their efficiency, which in turn predicted conceptual knowledge and procedural flexibility across the posttest and retention test. Clearly, effective comparisons are not limited to or better if examples share the same solution method. Rather, pairs of problem-solving examples can vary in problem features or solution methods, and contrasting solution methods seems particularly useful for supporting mathematics learning.

Contrasting solution methods is also more effective than sequential study of examples for supporting procedural knowledge and flexibility (Rittle-Johnson & Star, 2007). Thus, in two studies, comparing solution methods was most effective for supporting procedural flexibility, even on the more rigorous measure of flexibility used in the current study. Comparing solution methods seems particularly important for learning multiple procedures and when to use them across a variety of problem features. It also led to greater procedural knowledge, including greater accuracy on unfamiliar problem types, in Rittle-Johnson and Star (2007), but not in the current study. Rather, all three types of comparison were equally effective for supporting procedural knowledge, suggesting that comparison in general may be sufficient for learning procedures that can be adapted to novel problem features.

Unlike the current study, comparing solution methods was not more effective for supporting conceptual knowledge in Rittle-Johnson and Star (2007). It is difficult to interpret this difference in results across studies because the assessment was modified substantially. In Rittle-

Johnson and Star (2007), the assessment focused on general concepts of variables and

equivalence and had moderate internal consistency (α = .60). For the current study, we

modified the assessment to focus on the core concept highlighted in our instructional materials –

composite variables, and the measure had higher internal consistency (α = .74). Comparing

solution methods that do and do not capitalize on composite variable terms should help students

understand composite variables; a measure of conceptual knowledge that focuses on composite

variables may be critical for detecting the advantages of this condition given the brief and

focused nature of our intervention. Thus, we suspect that comparing solution methods is also

more beneficial for conceptual knowledge than sequential study of examples, but this hypothesis

warrants additional research.

Next, we consider the implications of our research on comparison for identifying

dimensions of comparisons that impact learning and for educational practice.

Compared to What? Dimensions of Comparison in Analogical Learning

Translating findings from cognitive science on comparison into an educational

intervention revealed a critical limitation in the literature. Prior research and theories of

analogical learning have focused almost exclusively on comparison of isomorphic examples –

examples with different surface features but the same solution or category structure. Clearly,

theories of analogical learning need to be expanded to consider how comparing different

solutions to the same problem supports learning. More generally, we consider the roles of

similarity and familiarity of examples in analogical learning.

Analogical learning occurs when two examples with some overlapping features are

aligned and inferences and abstractions are made based on this alignment (Gentner, 1983;

Hummel & Holyoak, 1997). Alignment requires one-to-one correspondence between features of

the examples, and correspondences are based on similarity in both the surface features and the underlying relations between objects in the examples.

Moderately similar, rather than highly similar, examples should help people ignore irrelevant surface features and abstract a more general underlying solution structure. Past research confirmed this prediction for category learning tasks (Tennyson et al., 1980; Tennyson & Tennyson, 1975; Waxman & Klibanoff, 2000). The current findings support this prediction for a problem solving task and extend it to outcomes not previously considered – procedural flexibility and conceptual knowledge. Examples that vary on one or a few important dimensions have been labeled contrasting examples, and contrasting examples may be particularly important when targeting outcomes beyond procedural knowledge (Bransford, Franks, Vye, & Sherwood, 1989).

Learning differences between the two moderately-similar comparison conditions indicate that overall similarity is an under-specified construct. For problem-solving tasks, two critical dimensions of examples that can vary are problem features and solution methods. The optimal feature to vary likely depends on the targeted domain and outcomes. In the case of equation solving, contrasting solution methods better facilitated procedural flexibility as well as conceptual knowledge.

In addition, the optimal similarity and dimension of contrast may depend on learners' familiarity with the domain. In the current study, the students had learned to solve one and two-step equations in prior lessons, and thus were not novice equation solvers. Most children knew the conventional, distribute-first method for solving equations at pretest; 68% of students used this method at least once at pretest. Prior research on category learning suggests that novices in a domain sometimes do not learn from comparing moderately similar examples and that high

similarity may be needed early in the learning process (Gentner & Namy, 2004). Thus, moderate similarity in examples may only be optimal for learners with some prior knowledge in the domain.

In particular, it may be best if learners are somewhat familiar with one solution method before contrasting it with another, unfamiliar method. Although most participants were familiar with the conventional, distribute-first, method, few were familiar with the shortcut method that capitalized on composite variables. Only 10% of students ever used a shortcut method at pretest. Thus, students who compared methods were comparing a familiar method to a less familiar method. In the other two conditions, students compared two examples of a familiar method in some pairs and two examples of an unfamiliar method in other pairs. According to Gentner (2005), there are two kinds of analogical learning based on whether one example is familiar or neither example is familiar (projective analogy vs. analogical encoding). Previous research has not evaluated whether or when one type of analogical learning is preferable to the other. It may be that analogies from a familiar example to an unfamiliar example support greater learning than analogies between two unfamiliar examples. If so, it may be important to ensure familiarity with one type of example before introducing the other. For example, it may be best for students with little prior knowledge of a task to first become familiar with one solution method through studying examples individually (or through comparing equivalent examples of that solution method) and then go on to compare the familiar method to an alternative solution method. This is an important direction for future research because of its theoretical implications for the interaction between familiarity and similarity of examples and for its educational implications for lesson sequencing.

Implications for Reform Efforts in Mathematics Education

The current findings provide much needed evidence in support of reform efforts in mathematics education that advocate for comparison of solution methods. Our unique use of random assignment of students to condition within their regular classroom context, along with maintenance of a fairly typical classroom environment, provided causal evidence for the benefit of comparing solution methods while maintaining fairly good external validity. Comparing solution methods may indeed be more effective than other types of comparison for supporting mathematical learning across a variety of measures. Our findings also show benefits for comparing solution methods in a more diverse set of students and teachers than in our previous study (Rittle-Johnson & Star, 2007). The findings provide at least three general suggestions for using comparison in mathematics classrooms.

First, comparing two meaningfully different solution methods seems most beneficial to learning. However, if not carefully orchestrated in classrooms, teachers may not chose the appropriate solution methods or problems for comparison to be effective. As noted in the introduction, US teachers commonly use comparison in their lessons, but frequently *not* in ways that seem most conducive to the development of mathematical understanding (Richland et al., 2004; Richland et al., 2007). In the current context, teachers may err by having students compare two *trivially* different solutions to the same problem (e.g., for the problem $3x + 2 = 5x + 7$, comparing subtracting $3x$ from both sides first versus subtracting 2 from both sides first). Such a comparison is more similar to the *equivalent* condition, where the solution methods only differed on superficial features. To implement comparisons that are more similar to our *methods* condition, teachers must choose problems and solution methods carefully. The problems should highlight important and meaningful concepts for students to learn, such as our choice to focus on

problems with composite variables. These problems need to have multiple solutions, preferably ones that differentially capitalize on important problem features. At the same time, the methods may need to be sufficiently similar for students to be able to align the methods and make meaningful comparisons. Finally, students may need some familiarity with one of the methods before comparing two different methods. Simply suggesting that teachers have students compare alternative solution methods is unlikely to be sufficient for improving teaching and learning.

Second, comparing problem types may be a good use of comparison when the targeted learning outcome is flexible use of instructed solution methods or when meaningfully different solutions are not prevalent. Comparing problem types was as effective as comparing solution methods at promoting use of a novel solution method. Contrasting examples of how a novel solution method changes and stays the same when important problem features vary can facilitate learning of that method (Cummins, 1992; Gick & Paterson, 1992; VanderStoep & Seifert, 1993). Often, the goal in mathematics is for students to learn a new solution method and apply it broadly; comparing problem types may help broaden the use of the new method.

Third, comparison requires careful support to be effective (Richland et al., 2007). Our materials were carefully designed to support effective comparison. Past research suggests that five features of our intervention may have been particularly important. As noted in Rittle-Johnson and Star (2007), three of these features are 1) a written record of all to-be-compared solution methods, with the solution steps aligned (Fraivillig, Murphy, & Fuson, 1999; Richland et al., 2004; Richland et al., 2007) 2) explicit opportunities to identify similarities and differences in methods (Catrambone & Holyoak, 1989; Fraivillig et al., 1999; Gentner et al., 2003; Lampert, 1990; Silver et al., 2005) and 3) instructional prompts to encourage students to consider the efficiency of the methods (Fraivillig et al., 1999; Lampert, 1990). Two additional features were

using common labels, such as labeling the solution steps, to invite comparison and help

alignment (Namy & Gentner, 2002) and providing some direct instruction to supplement

learners' comparisons (Schwartz & Bransford, 1998; Tennyson & Tennyson, 1975; VanderStoep

& Seifert, 1993). In the current study, scaffolds for effective comparison were embedded in the

instructional materials. Worked examples, carefully crafted explanation prompts, and peer

collaboration seemed to support productive explanation during partner work in the classroom and

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. At the same time, we caution that poorly planned or implemented

comparison is unlikely to facilitate learning.

Limitations and Future Directions

Before advocating for widespread use of comparison in mathematics instruction, it is

critical to evaluate the benefits and drawbacks of comparison under more typical classroom

conditions. For example, many of the students taking pre-algebra in middle school are

considered advanced in mathematics, so it is important to evaluate the effects of comparison with

students with lower mathematical ability. In addition, the instruction in this study was

implemented largely by researchers, rather than classroom teachers. How effectively classroom

teachers would implement comparison is important to evaluate. It is also important to evaluate

the effects of comparison with a variety of mathematical topics and on standardized tests. We

have found that comparing solution methods is more effective than sequential study of the same

examples for fifth graders learning about computational estimation (Star & Rittle-Johnson,

2008), but we have not yet explored different types of comparison in different domains.

In addition to these practical issues, the current study raises new theoretical questions on what features of examples to compare and how these different types of comparisons impact learning. For example, how important are similarity and familiarity of the examples?   Do these two factors interact and are the effects consistent across learners?

In conclusion, comparison seems to be a fundamental learning process, but it matters what is being compared. Comparing multiple solution methods to the same problem facilitated learning, particularly conceptual knowledge and procedural flexibility. Contrasting examples, rather than highly similar ones, can help optimize learning on a variety of measures. Future research and educational practice must pay careful attention to what is being compared.

References

Ball, D. L. (1993). With an Eye on the Mathematical Horizon: Dilemmas of Teaching

Elementary School Mathematics. *The Elementary School Journal, 93*, 373-397.

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

Dialogues, 3,* 11.

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*(1), 87-

106.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy

of educational objectives. In B. S. Bloom (Ed.), *Taxonomy of educational objectives: The

classification of educational goals by a committee of college and university examiners*

(pp. 186-193). New York: Longman, Greens.

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.

Bransford, J. D., Franks, J. J., Vye, N. J., & Sherwood, R. D. (1989). New approaches to

instruction: Because wisdom can't be told. In S. Vosniadou & A. Ortony (Eds.),

Similarity and analogical reasoning (pp. 470-497). New York, NY: Cambridge

University Press.

Carpenter, T. P., Franke, M. L., Jacobs, V. R., Fennema, E., & Empson, S. B. (1998). A

    longitudinal study of invention and understanding in children's multidigit addition and

    subtraction. *Journal for Research in Mathematics Education, 29*(1), 3-20.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-

    solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition,*

    15(6), 1147-1156.

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

    19, 2-10.

Cummins, D. (1992). Role of analogical reasoning in the induction of problem categories.

    Journal of Experimental Psychology: Learning, Memory, and Cognition, 18(5), 1103-

    1124.

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.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science,*

    7(2), 155-170.

Gentner, D. (2005). The Development of Relational Category Knowledge. In D. H. Rakison & L.

    Gershkoff-Stowe (Eds.), *Building object categories in developmental time* (pp. 245-275).

    Mahwah, NH: Lawrence Erlbaum.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for

    analogical encoding. *Journal of Educational Psychology, 95*(2), 393-405.

Gentner, D., & Namy, L. L. (2004). The Role of Comparison in Children's Early Word Learning. In S. R. Waxman & D. G. Hall (Eds.), *Weaving a lexicon* (pp. 533-568). Cambridge, MA: MIT Press.

Gick, M. L., & Holyoak, K. J. (1980). Analogical Problem Solving. *Cognitive Psychology, 12*(3), 306-355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1-38.

Gick, M. L., & Paterson, K. (1992). Do contrasting examples facilitate schema acquision and analogical transfer? *Canadian Journal of Psychology, 46*(4), 539-550.

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

Hiebert, J., & Wearne, D. (1996). Instruction, Understanding, and Skill in Multidigit Addition and Subtraction. *Cognition and Instruction, 14*(3), 251-283.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104*(3), 427-466.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford Press.

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

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.

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*(1), 5-15.

National Council of Teachers of Mathematics. (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, Va: NCTM.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. Reston, Va: National Council for Teachers of Mathematics.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for Prekindergarten through Grade 8 Mathematics*. Reston, Va: National Council for Teachers of Mathematics.

National Mathematics Advisory Panel. (2008). *Foundations of Success: The Final Report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy, 7*(1), 85-98.

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122-133.

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*(4), 525-556.

Reed, S. K. (1989). Constraints on the abstraction of solutions. *Journal of Educational Psychology, 81*(4), 532-540.

Reed, S. K., Ackinclose, C. C., & Voss, A. A. (1990). Selecting analogous problems: Similarity versus inclusiveness. *Memory and Cognition, 18*(1), 83-98.

Richland, L. E., Holyoak, K. J., & Stigler, J. W. (2004). Analogy Use In Eighth-Grade Mathematics Classrooms. *Cognitiion and Instruction, 22*, 37-60.

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

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91*(1), 175-189.

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*(2), 346-362.

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561-574.

Ross, B. H., & Kennedy, P. K. (1990). Generalizing From the Use of Earlier Examples in Problem Solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 42-55.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*(2), 147-177.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*(4), 475-522.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

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.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

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

Star, J. R., & Rittle-Johnson, B. (2007). Flexibility in problem solving: The case of equation solving. *Learning and Instruction, doi:10.1016/j.learninstruc.2007.09.018*.

Star, J. R., & Rittle-Johnson, B. (2008). *The Role of Comparison in the Development of Flexible Knowledge of Computational Estimation. .* Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York.

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

Tennyson, C. L., Tennyson, R. D., & Rothen, W. (1980). Content structure and instructional control strategies as design variables in concept acquisition. *Journal of Educational Psychology, 72*(4), 499-505.

Tennyson, R. D., & Tennyson, C. L. (1975). Rule acquisition design strategy variables: Degree of instance, divergence, sequence, and instance analysis. *Journal of Educational Psychology, 67*(6), 852-859.

VanderStoep, S. W., & Seifert, C. M. (1993). Learning "how" versus learning "when": Improving transfer of problem-solving principles. *Journal of the Learning Sciences, 3*(1), 93-111.

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47*, 513-539.

Waxman, S. R., & Klibanoff, R. S. (2000). The role of comparison in the extension of novel adjectives. *Developmental Psychology, 36*(5), 571-581.

Footnote

[1] Given that students discussed their explanations, it is not surprising that intraclass correlations between partners on these measures was high. Therefore, we used multi-level modeling. Our model had two levels – the individual level and the dyad level. Effects of experimental condition were tested in the second-stage (dyad-level) analyses. We specified the use of restricted maximum likelihood (REML) estimation and compound symmetry for the variance-covariance structure in the models (Kenny et al., 2006).

Table 1

How Features Of The Example Pairs Varied Across Conditions

Compare Condition	Problem	Solution Method	Similarity Rating on 9 pt scale [a]
Equivalent	Same Type	Same	8.36
Problem Types	Different Type	Same	5.48
Solution Methods	Same Problem	Different	5.27

[a]To verify the similarity of example pairs in the 3 conditions, we had a group of 26 undergraduates rate the similarity of each of the worked-example pairs used across the conditions on a scale from 1 (highly dissimilar) to 9 (highly similar). Undergraduates rated the equivalent example pairs as highly similar and significantly more similar than the problem types or solution methods pairs, $F(1, 24) = 156.6$, $p < .001$, which did not differ from each other.

Table 2

Alternative Solution Methods for The Three Types of Equations

Equation Type[a]	Sample Solution via Conventional, Distribute-First Method	Sample Solution via Composite-Variable Shortcut Method
$a(x + b) = c$	$3(x + 1) = 15$	$3(x + 1) = 15$
Divide composite	$3x + 3 = 15$	$x + 1 = 5$
	$3x = 12$	$x = 4$
	$x = 4$	
$a(x + b) + d(x + b) = c$	$2(x + 1) + 3(x + 1) = 10$	$2(x + 1) + 3(x + 1) = 10$
Combine composite	$2x + 2 + 3x + 3 = 10$	$5(x + 1) = 10$
	$5x + 5 = 10$	$x + 1 = 2$
	$5x = 5$	$x = 1$
	$x = 1$	
$a(x + b) = d(x + b) + c$	$7(x - 2) = 3(x - 2) + 16$	$7(x - 2) = 3(x - 2) + 16$
Subtract composite	$7x - 14 = 3x - 6 + 16$	$4(x - 2) = 16$
	$7x - 14 = 3x + 10$	$x - 2 = 4$
	$4x - 14 = 10$	$x = 6$
	$4x = 24$	
	$x = 6$	

[a] "x" stands for a variable and other letters were replaced with numbers.

Table 3

Sample Items for Assessing Problem Solving, Flexibility, and Conceptual Understanding

Problem Type	Sample Items	Scoring and Scale Alpha
I. Conceptual Knowledge		$\alpha = .74$
(n = 9)	1. Here are two equations: $$98 = 21x$$ $$98 + 2(x + 1) = 21x + 2(x + 1)$$ (a) Look at this pair of equations. Without solving the equations, decide if these equations are equivalent (have the same answer). (b) Explain your reasoning.	1a. 1 pt for selecting "YES (have same answer)" 1b. 1 pt for noting equivalent because same thing added to both sides or cancel out.
	2. Which of the following is a like term to (could be combined with) $7(j + 4)$? (a) $7(j + 10)$, (b) $7(p + 4)$, (c) j, (d) $2(j + 4)$, (e) a and d	2. 1 pt for choice d
II. Procedural Knowledge		$\alpha = .73$
a. Familiar (n = 3)	$$1/2\ (x - 1) = 10$$ $$3(h + 2) + 4(h + 2) = 35$$	1 pt for each correct answer
b. Near Transfer (n = 3)	$$\frac{3m - 2}{5} = \frac{7}{5}$$ $$3(2x + 3x - 4) + 5(2x + 3x - 4) = 48$$	1 pt for each correct answer
c. Far Transfer (n = 2)	Solve the equation for h: $P = 2(b + h)$	Omitted; No student got correct at pre- or posttest
d. Mental Math (n = 2)	Solve this equation in your head after seeing it for 30 sec.: $3(x + 1) + 2(x + 1) = 10$	1 pt for each correct answer

III. Procedural Flexibility

a. Flexible Use	Use of more efficient composite-variable method on familiar and near transfer problems	$\alpha = .78$ see Table 5
b. Flexibility Knowledge		$\alpha = .81$
i. Generate multiple methods (n = 6)	Solve this equation in three different ways: $0.5(d + 3) = 10$	1 pt for two correct, unique solutions; 2 pts for three correct, unique solutions
ii. Recognize multiple first steps (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
iii. Evaluate non-conventional methods (n = 2)	$5(x + 3) + 6 = 5(x + 3) + 2x$ $6 = 2x$ a. What step did the student use to get from the first line to the second line? b. Do you think that this is a good way to start this problem? Circle One: (a) a very good way; (b) OK to do, but not a very good way; (c) Not OK to do b explain. Explain your reasoning. c. For which of the following equations would it be good to use the student's way to start the problem? (a) $6(x + 4) + 20 = 8(x + 4)$; (b) $10x = 11(x + 1)$; (c) $15(y + 23) + 40 = 16(y + 30)$; (d) none of the above	Part a: 1 pt for correctly identifying step. Part b: 2 pt for choice a, 1 pt for choice b. Part b explain: 2 pt if note efficiency or justify why ok; 1 pt if vague or prefer alternative step; 0 pt if incorrect content or blank Part c: 1 pt for choice a

Note: Cronbach's alpha is reported as the average across posttest and retention test. Alpha was lower on pretest (α = .59 to .70) because students were at floor on many items.

Table 4

Student Accuracy By Condition (Percent Correct)

	Pretest		Posttest		Retention Test	
	M	*SD*	*M*	*SD*	*M*	*SD*
Conceptual Knowledge						
Methods	29.9	21.7	52.8	25.4	59.7	28.3
Problem types	31.0	19.9	47.7	25.2	52.3	24.5
Equivalent	33.2	23.0	47.2	28.0	54.6	31.0
Procedural Knowledge						
Methods	22.0	20.6	53.81	29.4	64.5	27.4
Problem types	19.4	18.1	46.76	30.4	62.2	26.5
Equivalent	22.0	20.6	53.81	29.4	64.5	27.4
Flexibility Knowledge						
Methods	27.5	14.8	55.12	18.8	62.1	22.2
Problem types	29.5	12.2	52.77	15.6	58.2	19.2
Equivalent	33.8	14.7	52.82	17.9	57.2	19.7

Table 5

Solution Method by Condition (Percentage of Trials)

	Distribute First [a]	Composite-Variable Shortcut [a]	Other Algebra [b]	Non-Algebra	Blank
Pretest					
Methods	43.7	2.0	24.4	12.1	18.3
Problem types	41.6	1.2	17.1	12.8	27.7
Equivalent	43.0	3.9	16.3	9.9	27.6
Posttest					
Methods	30.6	39.8	19.1	1.5	9.0
Problem types	33.8	43.3	11.2	0.5	11.3
Equivalent	41.2	33.6	13.4	2.1	9.7
Retention Test					
Methods	24.8	52.0	14.7	1.3	7.2
Problem types	31.2	48.3	9.7	1.3	9.4
Equivalent	39.4	40.0	12.3	0.7	7.8

[a] *Equivalent* differs from other two conditions across posttest and retention test at p's $< .05$;

[b] *Problem types* differs from *Methods* at $p = .014$.

Table 6

ANCOVA results for Learning Outcomes

Outcome	Conceptual Knowledge			Procedural Knowledge			Flexibility: Use of Shortcuts			Flexibility Knowledge		
Factor	F-value[a]	*p*	η^2	F-value	*p*	η^2	F-value	*p*	η^2	F (1, 153)	*p*	η^2
Condition	5.76	.004	.07	0.92	.402	.01	4.96	.008	.06	5.01	.008	.06
Concept Pre	33.32	<.001	.18	4.73	.031	.03	10.32	.002	.06	5.09	.025	.03
Procedure Pre	5.36	.022	.03	11.44	.001	.07	1.06	.306	.01	12.85	<.001	.08
Flexibility Pre	8.31	.005	.05	6.68	.011	.04	5.57	.020	.04	14.31	<.001	.09
Ability Group	16.20	<.001	.10	10.97	.001	.07	3.99	.048	.03	26.90	<.001	.15
School	4.15	.018	.05	7.65	.001	.09	1.88	.156	.02	4.30	.015	.05
Time	7.90	.006	.05	23.19	<.001	.13	6.23	.014	.04	22.02	<.001	.13

[a]Condition and School degrees of freedom are (2, 153), all others are (1, 153).

Table 7

Percentage of Intervention Explanations Containing Comparisons, by Condition

Explanation Characteristic	Sample Explanations	Methods	Problem-Types	Equivalent
Compare efficiency [a]	"Jill used more steps" "She skipped an unneeded step"	21	9	1
Compare solution steps [b]	"Jessica distributed and Mary combined like terms" "They both multiplied"	35	46	32
Compare problem features [c]	"They both have like terms" "They have different numbers"	16	17	34
Compare answers [d]	"They end up with the same answer after all the steps"	8	2	5
Any Comparison	At least one comparison	72	76	69

[a] All three conditions different from each other at $p < .001$; [b] *Problem Types* differ from other two conditions at p's $< .001$. [c] *Equivalent* differs from other two conditions at $p < .001$; [d] *Method* differs from *Problem Types* at $p = .002$.

Table 8

Percentage of Intervention Explanations Containing Each Feature, By Condition

Explanation Characteristic	Sample Explanations	Methods	Problem-Types	Equivalent
1. References multiple methods [a]	"It is okay to do it either way."	69	62	43
2. Focus				
on method [a]	"He divided each side by 2."	83	82	64
shortcut	"Mary combined like terms."	28	24	31
on problem [a]	"There are different variables."	17	19	39
on answer [b]	"The answer is right."	9	3	10
3. Evaluates				
Efficiency [a, b]	"James' way was just faster."	37	17	7
Accuracy [a]	"Sammy's solution is also correct"	2	2	8
Problem Features [c]	"Heather's problem has easier number"	5	11	16
No Judgment [c]	"It's the opposite of multiplying"	52	65	68
4. Justify Mathematically [d]	"Used the right properties at the right times."	11	15	21

[a] *Equivalent* differs from other two conditions at p < .001; [b] *Problem types* differs from other two conditions at p < .002; [c] *Methods* differs from other two conditions at p < .001; [d] *Methods* differs from *Equivalent* at p < .001

Figure Captions

Figure 1. Sample packet page for each condition.

Figure 2. Conceptual knowledge by condition. (Estimated marginal mean across posttest and retention test. Error bars are standard errors.)

Figure 3. Procedural flexibility based on a) use of composite-variable shortcut methods and b) knowledge of multiple and efficient solution methods. (Estimated marginal means across posttest and retention test. Error bars are standard errors.)

Figure 1

A. Compare Solution Methods

Nathan's Solution:	Patrick's Solution:

Nathan's Solution:

$$5(y + 1) = 3(y + 1) + 8$$
$$5y + 5 = 3y + 3 + 8 \qquad \textit{Distribute} \underline{\quad\quad}$$
$$5y + 5 = 3y + 11 \qquad \textit{Combine} \underline{\quad\quad}$$
$$2y + 5 = 11 \qquad \textit{Subtract} \underline{\quad\quad} \textit{on Both}$$
$$2y = 6 \qquad \textit{Subtract} \underline{\quad\quad} \textit{on Both}$$
$$y = 3 \qquad \textit{Divide} \underline{\quad\quad} \textit{on Both}$$

Patrick's Solution:

$$5(y + 1) = 3(y + 1) + 8$$
$$2(y + 1) = 8 \qquad \textit{Subtract } \underline{3(y + 1)} \textit{ on Both}$$
$$y + 1 = 4 \qquad \textit{Divide} \underline{\quad\quad} \textit{on Both}$$
$$y = 3 \qquad \textit{Subtract} \underline{\quad\quad} \textit{on Both}$$

5. Describe 2 ways that these students' solutions are **different.**

6. What must be true about an equation for Patricks' way to be easier than Nathan's way?

B. Compare Problem Types

Abby's Solution:	Patrick's Solution:

Abby's Solution:

$$3(h - 2) + 5(h - 2) = 24$$
$$8(h - 2) = 24 \qquad \textit{Combine} \underline{\quad\quad}$$
$$h - 2 = 3 \qquad \textit{Divide} \underline{\quad\quad} \textit{on Both}$$
$$h = 5 \qquad \textit{Add} \underline{\quad\quad} \textit{on Both}$$

Patrick's Solution:

$$5(y + 1) = 3(y + 1) + 8$$
$$2(y + 1) = 8 \qquad \textit{Subtract } \underline{3(y + 1)} \textit{ on Both}$$
$$y + 1 = 4 \qquad \textit{Divide} \underline{\quad\quad} \textit{on Both}$$
$$y = 3 \qquad \textit{Subtract} \underline{\quad\quad} \textit{on Both}$$

5. Describe one way the students' **problems** are the same and one way they are different.

6. Patrick's first step is **different** from Abby's first step because:

C. Compare Equivalent

Patrick's Solution:	Flynn's Solution:

Patrick's Solution:

$$5(y + 1) = 3(y + 1) + 8$$
$$2(y + 1) = 8 \qquad \textit{Subtract } \underline{3(y + 1)} \textit{ on Both}$$
$$y + 1 = 4 \qquad \textit{Divide} \underline{\quad\quad} \textit{on Both}$$
$$y = 3 \qquad \textit{Subtract} \underline{\quad\quad} \textit{on Both}$$

Flynn's Solution:

$$10(x + 3) = 6(x + 3) + 16$$
$$4(x + 3) = 16 \qquad \textit{Subtract} \underline{\quad\quad} \textit{on Both}$$
$$x + 3 = 4 \qquad \textit{Divide} \underline{\quad\quad} \textit{on Both}$$
$$x = 1 \qquad \textit{Subtract} \underline{\quad\quad} \textit{on Both}$$

5. Describe one way the students' **problems** are the same and one way they are different.

6. Flynn's first step is **similar** to Patrick's first step because:
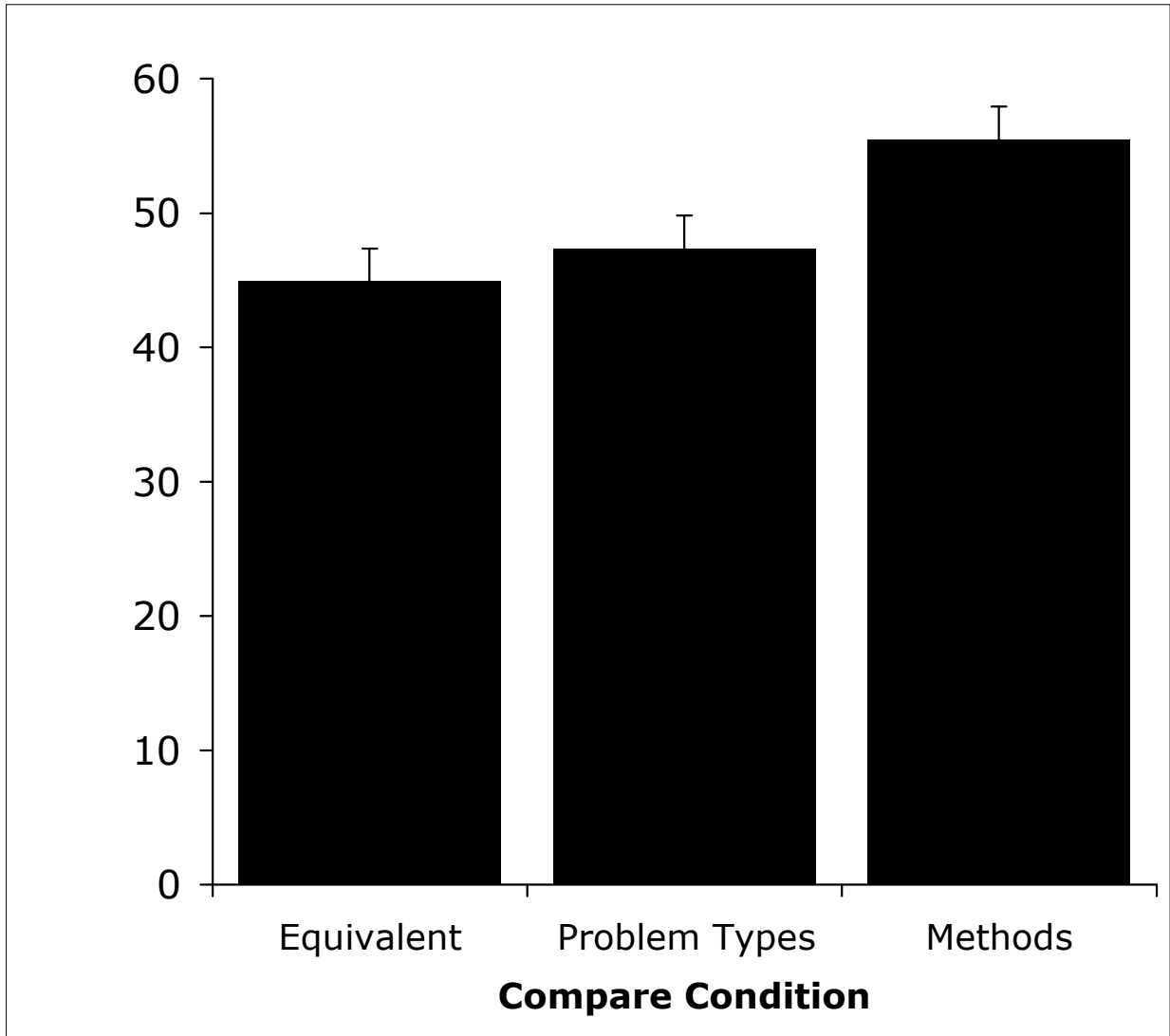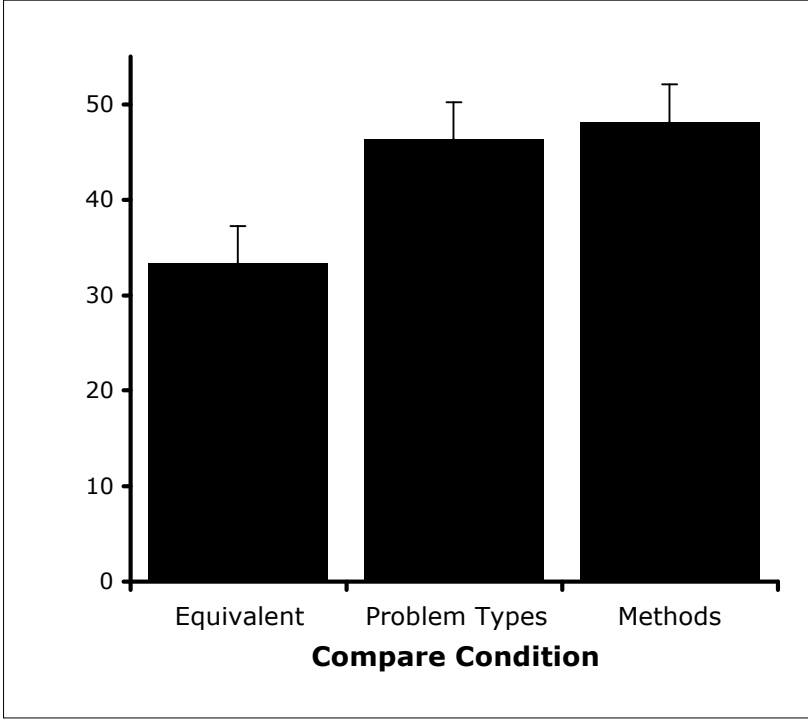
Figure 2

Figure 3
a)

b)