THE JOURNAL OF EXPERIMENTAL EDUCATION
2016, VOL. 0, NO. 0, 1–25
http://dx.doi.org/10.1080/00220973.2016.1180277

Routledge
Taylor & Francis Group

LEARNING, INSTRUCTION, AND COGNITION

Embellishing Problem-Solving Examples with Deep Structure Information Facilitates Transfer

Hee Seung Lee[a], Shawn Betts[b], and John R. Anderson[b]

[a]Yonsei University, Seoul, Korea; [b]Carnegie Mellon University, Pittsburgh, PA

ABSTRACT

Appreciation of problem structure is critical to successful learning. Two experiments investigated effective ways of communicating problem structure in a computer-based learning environment and tested whether verbal instruction is necessary to specify solution steps, when deep structure is already embellished by instructional examples. Participants learned to solve algebra-like problems and then solved transfer problems that required adjustment of learned procedures. Experiment 1 demonstrated that verbal instruction helped learning by reducing learners' floundering, but its positive effect disappeared in the transfer. More importantly, students transferred better when they studied with examples that emphasized problem structure rather than solution procedure. Experiment 2 showed that verbal instruction was not necessarily more effective than nonverbal scaffolding to convey problem structure. Final understanding was determined by transparency of problem structure regardless of presence of verbal instruction. However, verbal instruction had a positive impact on learners by having them persist through the task, and optimal instructional choices were likely to differ depending on populations of learners.

KEYWORDS

Instructional example; mathematical problem solving; nonverbal scaffolding; problem structure; verbal instruction

MANY TIMES WE learn to solve problems by studying examples of problem solution without verbal guidance. Sometimes this is by necessity because an example is all that is available (e.g., Lewis, 1988), and sometimes it is by choice because the student finds the expository text too difficult (e.g., Pirolli & Anderson, 1985). However, example solutions often have critical features that are not salient. When students fail to understand such features, they may memorize solution steps, resulting in poor transfer performance (Hiebert & Lefevre, 1986; Skemp, 1976). For example, in understanding the solved problem of $5x - x = 4x$, the student has to appreciate that $5x - x = 5^{*}x - 1^{*}x$ and that this can be factored thus, $(5 - 1)^{*}x$. If such critical structure is not presented students may form superficial rules. For instance, in this example a student may infer the superficial rule of decrementing the integer, and faced with $-5x - x$, may produce $-4x$. In contrast, when the underlying structure is apparent in a problem-solving example, students may not need explanations about the problem solution. Instead, students may be able to infer the correct rules just by applying their current knowledge to the presented example.

This research aimed to investigate the effect of making problem structure apparent on learning and transfer in mathematical problem solving. We also tested whether provision of verbal instruction is really necessary to specify the solution when the problem structure is already made apparent by instructional examples. We tested this with two different populations of learners in order to see whether our instructional principles, found from a lab-based study with college students, can be

CONTACT Hee Seung Lee hslee00@yonsei.ac.kr Department of Education, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Korea.

Color versions of one or more of the figures in this article can be found online at www.tandfonline.com/vjxe.

generalized to other populations and whether appropriate instructional supports differ depending on different populations of learners. To gain more control over background knowledge, we used isomorphs of algebra problems so that participants would all start out without knowing how to solve problems. Then we used algebra to embellish problem-solving examples with information about their underlying structure.

Effects of instructional explanations added to worked examples

There is strong evidence that learning is facilitated by the provision of worked examples (e.g., Atkinson, Derry, Renkl, & Wortham, 2000; Carroll, 1994; Renkl, 2002, 2005, 2011; Sweller & Cooper, 1985; Trafton & Reiser, 1993; Tuovinen & Sweller, 1999). Worked examples typically consist of a problem statement, solution steps, and a final answer to the problem (Atkinson et al., 2000; Renkl, Stark, Gruber, & Mandl, 1998; Wittwer & Renkl, 2010), and they are known to facilitate learning by reducing search activity that is irrelevant for problem schema acquisition and instead helping students to focus on relevant solution steps (Sweller, 2005; Sweller, Van Merriënboer, & Paas, 1998). Several studies also have shown benefits of adding worked examples within computer-based tutoring systems (e.g., Mathan, 2003; Salden, Aleven, Schwonke, & Renkl, 2010). However, there can be considerable variation in how much instructional guidance accompanies the worked example. For example, instructional explanations can be added to worked examples to explain underlying rules of problem solution steps or students can be left to induce a rule for themselves via self-explanations when making sense of each step of examples (Chi, 2000).

Providing instructional explanations seems to have both advantages and disadvantages and empirical evidence is mixed (e.g., Atkinson, 2002; Gerjets, Scheiter, & Catrambone, 2006; Große & Renkl, 2006; Lovett, 1992; Renkl, 2002; Ward & Sweller, 1990). The biggest advantage of providing instructional explanations is to guide students' understanding of the learning materials and avoid superficial induction (Renkl, 2002). When worked examples are provided, students sometimes fail to understand each step of the worked examples (Chi et al., 1989; Renkl, 1997; Renkl et al., 1998). If underlying solution steps are not explained and/or relevant features are not appropriately highlighted, students are left to generate their own explanations to understand the examples. This process is not always successful. Students often show *illusion of understanding* (Chi et al., 1989) and fail to solve comprehension problems without instructional explanations (Renkl, 2002).

Instructional explanations also have some disadvantages. Providing instructional explanations seems to reduce self-explanation activities and, in turn, learning outcomes (Schworm & Renkl, 2006). It can also impair learning by increasing extraneous cognitive load (split-attention effect, Tarmizi & Sweller, 1988; Mwangi & Sweller, 1998; Ward & Sweller, 1990). Ward and Sweller (1990) showed that when extra textual explanation was added to worked examples in an inappropriate way, positive effect of worked examples disappeared compared to conventional problem-solving practice conditions. Lastly, processing instructional explanations added to worked examples may increase study time. Gerjets et al. (2006) examined effect of amount of instruction added to worked examples on learning probability calculation problems. Although there was no effect of instructional amount on learning, students who received more instructional explanations spent more time studying than those who received fewer instructional explanations.

Effects of appreciating deep structure in problem solving

Several factors can contribute to the ineffectiveness of instructional explanations. One of the possible factors is whether instructional explanations focused on concepts and principles. Although all instructional explanations are designed with a purpose of teaching (Leinhardt, 2001; Leinhardt & Steele, 2005), different things can be emphasized in the explanations and this can affect what students pay attention to. Many theories of learning distinguish between conceptual and procedural knowledge (Bisanz & Lefevere, 1992; Rittle-Johnson & Alibali, 1999; Rittle-Jonson, Siegler, & Alibali, 2001) and students are required to learn both underlying concepts and procedures for correctly solving a problem in many domains. However, many students often fail to understand deep structure in a problem and simply focus on memorizing procedures or solution rules. Solution steps can be learned and practiced even without conceptual understanding, but learning without understanding can lead to poor transfer performance (Hiebert & Lefevre, 1986; Skemp, 1976).

For example, Perry (1991) showed that a principle-based instruction and a procedure-based instruction achieved comparable learning performance in teaching mathematical equivalence, but the principle-based instruction led to better transfer performance than the procedure-based instruction. Rittle-Johnson and Alibali (1999) also showed that conceptual instruction led to better transfer performance than procedure instruction in teaching mathematical equivalence concepts. Likewise, Ringenberg and VanLehn (2006) found that providing annotated worked examples was more efficient than providing procedure-based hints to obtain basic mastery in the domain of physics.

We think that the reported benefits of principle-based instructions occur when they identify the deep structure of a problem and so draw attention to relevant features that students might otherwise ignore. Many studies have shown that appreciation of problem structure is critical to successful problem solving. For example, analogical transfer is facilitated when deep structure of the example is highlighted rather than surface features. Structural feature refers to the underlying systems of relations between the elements, whereas surface feature refers to the objects or elements that are not causally relevant to goal attainment (Holyoak & Koh, 1987). For instance, in solving a word problem that involves computing total number of fruits, the name of fruits appearing in a problem (e.g., apples versus oranges) is a surface feature and thus does not affect problem solution, whereas the multiplication (e.g., number of apples per bag $\times$ number of bags) is a critical structural feature. The relevant structure can be highlighted by a verbal statement that emphasizes the structure (Gick & Holyoak, 1983), use of animated diagrams (Pedone, Hummel, & Holyoak, 2001), and comparison of multiple examples (Gick & Holyoak, 1983; Loewenstein, Thompson, & Gentner, 1999; Gentner, Loewenstein, & Thompson, 2003).

As another example, Jitendra and her colleagues have repeatedly shown benefits of instruction that emphasizes underlying mathematical structure in proportional reasoning (Jitendra et al., 2009; Jitendra et al., 2011; Jitendra, Star, Dupuis, & Rodriguez, 2013; Xin, Jitendra, & Deatline-Buchman, 2005). In the domain of arithmetic word problems, many other studies support benefits of explicit schema training (e.g., Fuchs et al., 2009; Fuchs et al., 2008; Fuchs et al., 2010; Fuson & Willis, 1989). Understanding of problem structure is also described as a defining characteristic of expert problem solvers (e.g., Chi, Feltovich, & Glaser, 1981; Schoenfeld & Hermann, 1982; Silver, 1981; Van Dooren, De Bock, Vleugels, & Verschaffel, 2010). In Chi et al.'s (1981) study, for example, experts were more likely to use structural information when categorizing physics problems, whereas novices relied on surface-level information.

Overall, past research demonstrates that instruction that focuses on problem structure is more effective. However, it is not still clear about how to communicate deep structure of a problem with learners. For example, does it have to take a verbal form of instruction? Can students learn without explicit verbal instruction as long as deep problem structure is highlighted? Answering these questions will shed light on why we have so many mixed results regarding effectiveness of instructional explanations. We propose that effectiveness of any type of instruction depends on whether the problem structure is made apparent to learners. Only with understanding of deep structure can students generate their own solution procedure correctly. If instructions ineffectively communicate deep problem structure, students will learn solution procedures based on surface feature of the problem. This will not necessarily lead to correct problem solving especially when problems are changed.

This research aims to investigate effective ways of communicating deep problem structure in a context of computer-based learning environment. Computer-based instructional systems, like intelligent tutoring systems, have attracted a lot of attention in education and are being widely used in classrooms; for example, Cognitive Tutor Algebra curriculum is being used in over 3,000 schools nationwide, and it has been successful in raising students' test scores (e.g., Koedinger, Anderson, Hadley, & Mark, 1997; Morgan & Ritter, 2002; Pane, Griffin, McCaffrey, & Karam, 2013). In this study, we wanted to create a computer-based learning environment that is similar to that of other intelligent tutoring systems, thus many of our instructional design choices (hints and feedback) were based on features of the Cognitive Algebra Tutor. Details of hint and feedback mechanisms will be described in the next section.

Two experiments investigated how to effectively communicate problem structure in a computer-based learning environment and tested whether verbal instruction is really necessary to specify solution steps when deep structure is already embellished by instructional examples. Prior research has reported mixed results regarding effectiveness of verbal instruction (e.g., Atkinson, 2002; Gerjets, Scheiter, & Catrambone, 2006; Große & Renkl, 2006; Lovett, 1992; Renkl, 2002; Ward & Sweller, 1990). As we proposed earlier, if effectiveness of verbal instruction depends on whether the problem structure is made apparent to learners, verbal instruction will help learning when the problem structure is not obvious to learners. In contrast, when learners can appreciate deep problem structure via other nonverbal features, verbal instruction will have little impact (if any) on learning. Experiment 1 tested this by examining relative effectiveness of verbal instruction when deep problem structure was highlighted versus not highlighted by (nonverbal) instruction examples. To manipulate the transparency of problem structure, we elaborated instructional examples with either the principle underlying problem solution (structural example) or the solution procedure only using nonverbal features (procedural example). A complication is that learners may have difficulty in processing such highlighted features. In that case, verbal instruction may help students avoid initial floundering as they try to interpret the highlighted information. If they can eventually interpret the material and achieve understanding they should not suffer a deficit in a later transfer phase. To test this, we included both learning and transfer sessions and examined whether effectiveness of instructional methods (verbal instruction and instructional examples) change depending on the type of study phases and test problems.

In Experiment 2 we tried to push our hypothesis further by putting structural information as well as procedural steps into the instructional explanations and compared this condition with a nonverbal scaffolded condition. Although previous studies have shown importance of highlighting problem structure (e.g., Gick & Holyoak, 1983; Gentner, Loewenstein, & Thompson, 2003; Jitendra et al., 2011; Perry, 1991), it is still unclear whether nonverbal scaffolding can be as effective as verbal explanations. Experiment 2 investigated relative effectiveness of verbal versus nonverbal instructional methods, and whether combining the two methods is more effective than adopting a single method. We tested this with college students and online participants to see whether our design choices could be generalized to other populations rather than specific college students. An instructional method that is effective with a certain group might not be effective with other groups (e.g., Kalyuga, 2007, Kalyuga et al., 2003; Salden et al., 2010). By testing two different populations we looked at how different types of instruction have impact on success of learning and how optimal instruction design can change.

Current experiments

We investigated the effects of embellishing problem-solving examples with problem structure and effects of providing verbal instruction on learning and transfer in the domain of mathematical problem solving. We wanted a learning task that would be like algebra but which would be novel to all participants. Therefore, we developed a computer-based instructional system,[1] like the one used by Lee et al. (2011), which was originally adapted from Brunstein, Betts, and Anderson (2009). In this system, participants solve a series of isomorphs of algebraic problems that are represented as data-flow diagrams. This system allowed us to study college students learning anew the equivalent of algebra. Besides the ease of using this population, the more important advantage of this system is that it allows us to study conditions that might not result in good learning without hurting the population's knowledge of any mathematics that would be used outside of the laboratory. Using data-flow diagrams, a sequence of learning units was constructed to correspond to the curriculum typically found in algebra courses. Participants first learned to solve problems equivalent to equations with a single appearance of a variable (e.g., $2 - x + 12 = 8$) and then learned to solve problems equivalent to equations with multiple appearances of a variable (we call this "linearize" problem, e.g., $(8 - x) + (5 * x) = 36$).

Figure 1(a) illustrates an example problem with one unknown variable (we call this "propagate" problem). An unknown number flows from the top box into the boxes below, the arithmetic operations are performed, and the final result is the *8* at the bottom. The participant's task is to determine what values to fill into the empty portions of the boxes. The diagram in Figure 1(a) is equivalent to the

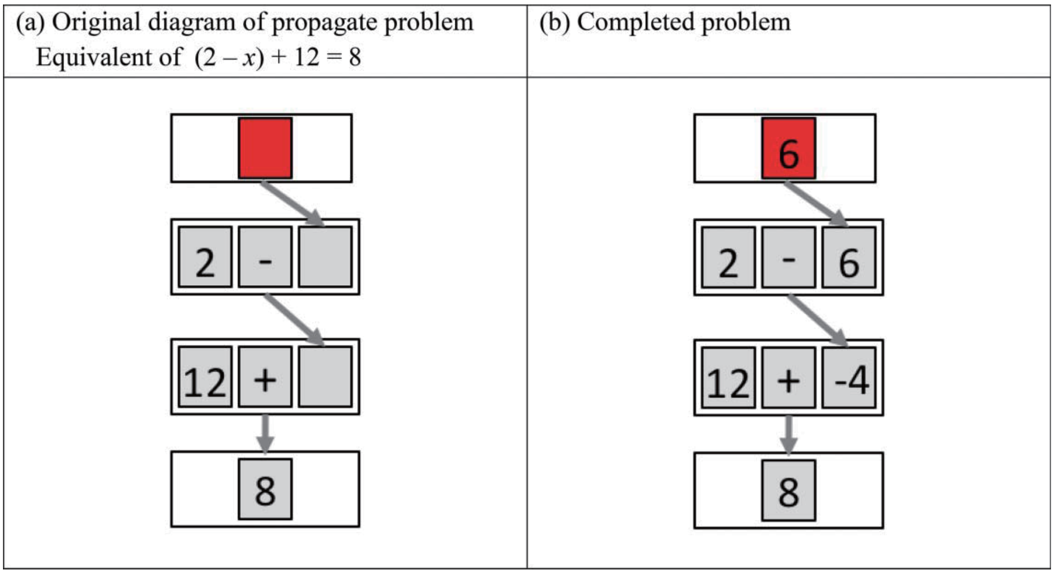


Figure 1. An example of a propagate problem. Propagate problems have a single appearance of a variable.

algebraic expression $(2-x) + 12 = 8$. In this problem the unknown value can be determined simply by "propagating" the number up from the bottom, unwinding the arithmetic operations—placing -4 in the empty tile above the 8 (equivalent to rewriting $2 - x + 12 = 8$ as $2 - x = -4$), then placing *6* in the tile above it (equivalent to rewriting this as $x = 6$), and finally placing *6* in the top unknown box. Figure 1(b) shows the completed diagram in this propagate problem. In our current and other studies (e.g., Lee et al., 2011), we observed that most participants (both college-level and algebra-level students) found propagating numbers easy and intuitive thus did not need much assistance when learning to solve this kind of problem. A couple of participants described this as "like doing Sudoku."

When an unknown value flows down multiple paths, however, this simple propagation strategy no longer works and most participants find such problems dramatically more difficult. Figure 2 shows an example of such a problem (we call this "linearize" problem). This problem is essentially equivalent to solving an equation with multiple appearances of the variable. The diagram in Figure 2 is equivalent to the algebraic expression, $(8 - x) + (5 * x) = 36$, where an unknown value flows down into the boxes below and the result value becomes *36*. Different from propagate problems (e.g., Figure 1), in cases like

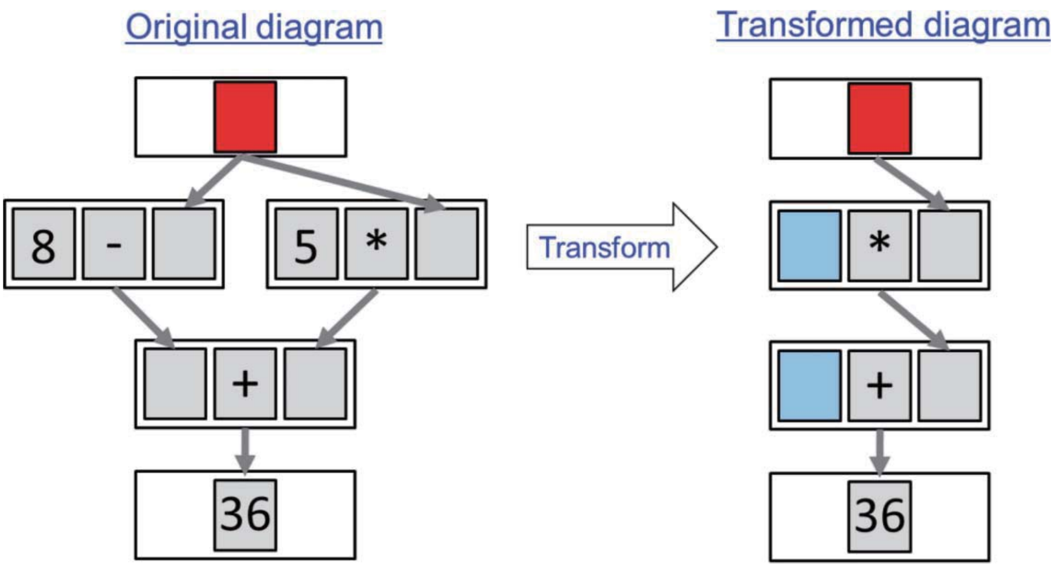


Figure 2. An example of a linearize (transform) problem used in the learning phase. The task is to find values to fill in the blue tiles of the transformed diagram. The correct answers for the two blue tiles are 4 and 8 because $(8 - x) + (5 * x) = 8 + 4x$. After filling in these two numbers, people can fill in the rest of the tiles by propagating numbers.

Figure 2, two paths converge in a single result, making a simple propagation procedure impossible. The way to solve this problem within the rules of the system is to transform the left graph into the form on the right and fill in the colored portions of the diagram with numbers. The problem requires understanding that the transformed diagram is a simplified version of original diagram.

To correctly solve the problem, students have to understand that the diagram transformation is equivalent to $(8 - x) + (5 * x) \rightarrow \mathrm{a} + \mathrm{b}x$ and then determine the values for a and b; that is, the transformed diagram has a structure that corresponds to the algebraic equivalent of the result of collecting like terms. However, this problem structure is not obvious in data-flow diagrams, and participants show difficulty in figuring out the values of a and b. In this example, the original diagram can be simplified into $(8 - x) + (5 * x) = 8 + 4x$, thus a is 8 and b is 4. The coefficient 4 goes to the box next to the "*" operator and the constant 8 goes into the box next to the "+" operator. After filling in these two values, the diagram becomes the linear form and now participants can apply the simple propagating strategy to fill in the rest of the empty tiles as in Figure 1.

Above we have tried to explain the solution of these problems by reference to algebraic transformations. In this study, we used algebraic interpretation of the data-flow diagrams as a way of embellishing problem-solving examples with their deep problem structure. Because propagate problems (Figure 1) are very easy, we focused on linearize problems (Figure 2) to apply experimental manipulations.

Experiment 1

In the first experiment, we developed two different kinds of instructional examples as a way of showing intermediate cognitive steps by either emphasizing the structure of a problem (structural example) or focusing on a proficient problem solver's procedural solution (procedural example). These two different types of examples were constructed such that the former explains "why" problems are solved in such a way, whereas the latter does not convey such information but rather focuses on the solution steps. This distinction is similar to Van Gog and her colleagues' (Van Gog, Paas, & Van Merriënboer, 2004, 2006, 2008) product-oriented versus process-oriented worked example. We expected that participants would be able to successfully learn from both kinds of example demonstrations but would develop different understanding, thus affecting transfer performance. Participants who studied with structural examples will be better able to adjust their learned procedure to deal with new cases. Crossed with our choice of example we manipulated whether there was accompanying verbal instruction, resulting in four different learning conditions.

Method

Participants

Ninety-six graduate and undergraduate students (43 male and 53 female, $M = 21$ years, $SD = 2.3$ years) from Carnegie Mellon University (CMU) participated in this study. Each participant was randomly assigned to one of the four conditions (24 structural/explanatory, 26 structural/nonexplanatory, 21 procedural/explanatory, and 25 procedural/nonexplanatory). Participants received \$10/hour plus a performance-based bonus. Two participants did not show up on the transfer phase and these two were excluded from the data analysis (1 structural/explanatory, 1 structural/nonexplanatory).

Design and materials

A 2×2 between-subjects design was employed to test the effects of different types of instructional examples and verbal instruction. First, two different types of examples (structural vs. procedural) were constructed to illustrate different aspects of the problem solution. These examples were available whenever participants requested a hint. Figure 3 shows an example of each instructional example condition. In the structural example condition, shown in Figure 3(a), when participants requested a hint, algebraic expressions were directly drawn onto the diagram participants were solving, to illustrate how the

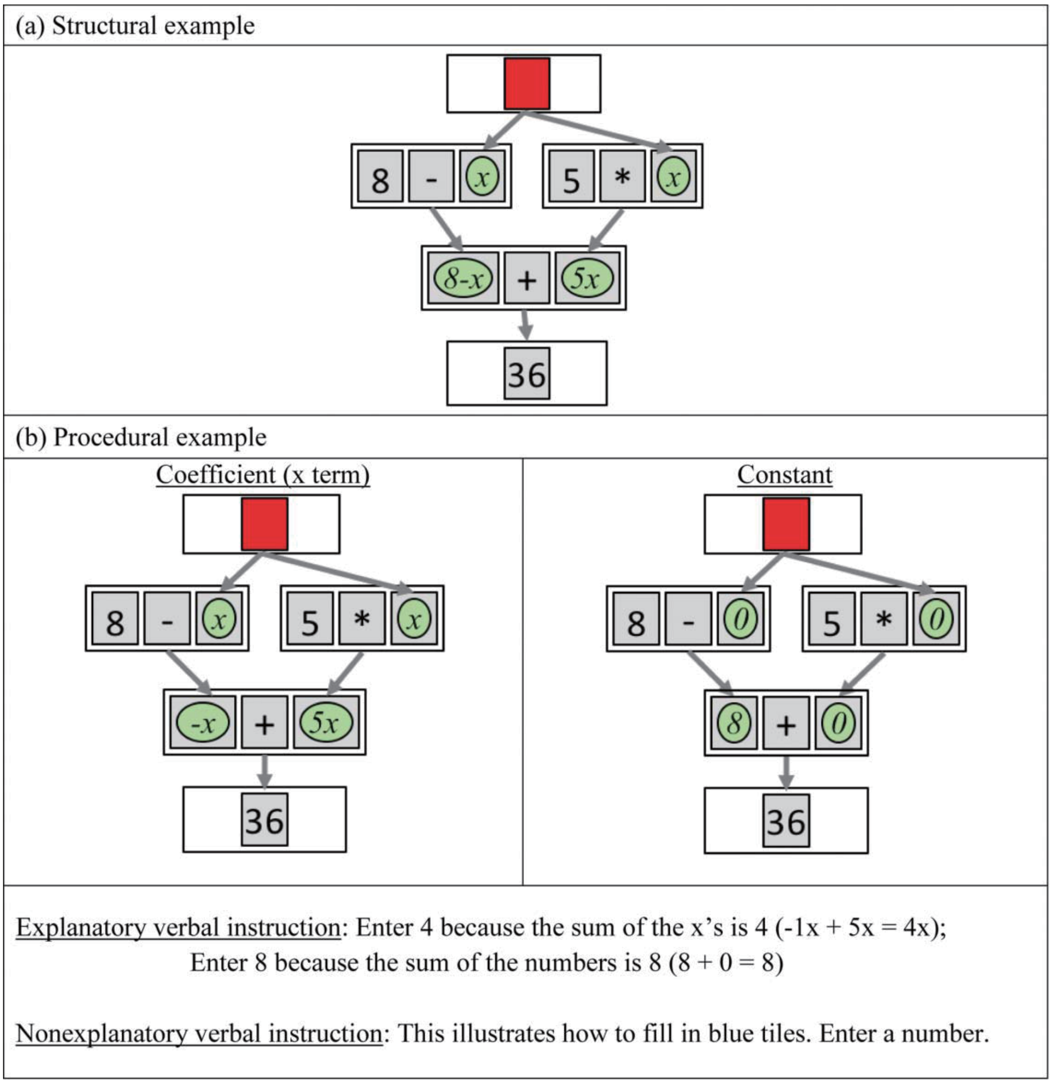


Figure 3. Example illustration for the (a) structural and the (b) procedural example condition. In the (a) structural condition, both coefficient and constant terms are computed simultaneously using equivalent algebraic expressions, whereas in the (b) procedural condition, solution steps for the coefficient term (left) are illustrated first and the constant term (right), second. Depending on the verbal instructional conditions, different hint texts were provided and example texts are shown below the figures for the explanatory and nonexplanatory conditions. Hint texts were identical for structural and procedural conditions. Depending on which box is being filled in, participants in the explanatory condition see the instruction for 4 or 8. In the nonexplanatory condition, participants see the same verbal information for both boxes.

diagram is equivalent to an algebraic expression. The example shows how the components of the diagram come together to be the equivalent of an algebraic expression on which one can perform collection of like terms. This type of example clearly shows that the data-flow diagram is just an equivalent of algebraic expression; thus, participants would have a basis for understanding the values that they are entering for coefficient and constant. However, if participants were to actually follow these steps mentally in solving a diagram, they would face large working-memory demands (especially when problems are large) because they would have to simultaneously compute and track intermediate results of both the coefficient and constant terms.

In the procedural example condition, illustrated in Figure 3(b), we illustrated the steps that proficient problem solvers in this system reported using in our previous studies (Lee et al., 2011; Lee et al., 2014; Lee et al., 2015). Proficient solvers, once after they understood that the data-flow diagram was equivalent to an algebraic expression, computed coefficient, and constant terms separately to reduce their computation load and increase accuracy. It was often hard to compute two terms simultaneously without paper and pencil, especially in large problems. Thus, we designed the procedural example to show the separate computations while not showing overall algebraic structure. When participants requested a hint, the instructional example was provided in two steps; the system first showed the calculation of the coefficient terms as in the left part of Figure 3(b) and then the calculation of the constant as in the right part of Figure 3(b). By focusing on just one component, working-memory load is reduced. Reducing cognitive load could possibly make computation more efficient and also enhance learning (Sweller, 2005; Sweller, Van Merriënboer, & Paas, 1998). However, a potential problem of teaching how to calculate the components in two steps is that the equivalent algebraic transformation is never illustrated. Rather, fragments are illustrated obscuring the algebraic understanding of the transformation.

Crossed with the structural versus procedural manipulation of instructional example, type of verbal instruction was manipulated by providing explanatory or nonexplanatory verbal instruction. In the explanatory condition, hint texts provided explanations of how intermediate answers were computed. In the nonexplanatory condition, the textual message just specified the actions required in the interface such as "click a box" and "enter a number." Thus in this condition, participants have to discover problem-solving rules on their own by analyzing the examples and the problem solutions. The nonexplanatory participants had to find answers by trial and error unless they were timed out (and the solution was provided without explanation). In contrast, in the explanatory conditions, verbal instruction provided answers for each step by participant request, enabling him or her to move forward even when the participant did not really understand problem solutions. This manipulation was applied during the entire learning phase. Example verbal instructions are shown below the diagrams in Figure 3 for the explanatory and nonexplanatory conditions.

In the transfer phase, participants were tested with problems that required transformation that they never encountered during the learning phase. According to Barnett and Ceci's (2002) taxonomy of transfer, our transfer task had high context similarity with the learning task including knowledge domain; modality; and physical, functional, and social context, except for temporal context (2 days of delay). Thus, our study mainly focused on content transfer—in particular, how well a learned procedure is accurately recalled and executed when there is a procedural change when solving a problem.

There were two types of transfer problems—graphic and algebraic problems. The graphic problems required dealing with a novel complexity in the diagram structure but the basic procedure for determining the answer was unchanged. In contrast, the algebraic problems required understanding of how the transformed diagram was equivalent to the original diagram and modifying the procedure accordingly. Table 1 shows the list of problem structure of graphic and algebraic problems (see Figure 4 for specific examples of graphic and algebraic problems). Graphic transfer problems have a more complex structure that requires parsing of graphic complexity, and participants have to figure out which boxes to include or exclude for computations. On the other hand, algebraic problems require an algebraic transformation after performing a collection of like terms. These two different types of transfer problems were constructed with different purposes. Graphic problems could be solved even if participants did not understand why the right and left graphs were equivalent but rather just understood how to combine numbers

Table 1. Examples of learning problems and transfer problems. Each example shows the type of required transformations. There was only one type of linearize problem in the learning phase and four subtypes of graphic and algebraic problems in the transfer phase. In Experiment 2, only algebraic problems were used in the transfer phase.

		Example problem and required transformation	Correct answers
Learning phase	Propagate	$2 - x + 12 = 8$ (No transformations required)	$x = 6$
	Linearize	$(8 - x) + (5^*x) \rightarrow a + bx$	$a = 8, b = 4$
Transfer phase	Graphic Problems		
	Type 1	$(((8 - x) + x) + x) + (5^*x) \rightarrow (a + bx) + x + (5^*x)$	$a = 8, b = 0$
	Type 2	$(((8 - x) + x) + x) + (5^*x) \rightarrow (a + bx) + (5^*x)$	$a = 8, b = 1$
	Type 3	$((8 - x) + 5) + (8 - x) \rightarrow a + bx'$ $(x' = 8 - x)$	$a = 5, b = 2$
	Type 4	$((8 - x) + 5) + (8 - x) \rightarrow a + bx$ $(x' = 8 - x)$	$a = 21, b = -2$
	Algebraic Problems		
	Type 1	$(8 - x) + (5^*x) \rightarrow (x + a)^*b$	$a = 2, b = 4$
	Type 2	$(8 - x) + (5^*x) \rightarrow$ -a $+ bx$	$a = -8, b = 4$
	Type 3	$(8 - x) + (5^*x) \rightarrow 4^*(a + bx)$	$a = 2, b = 1$
	Type 4 (Exp. 1)	$(8 - x) + (5^*x) \rightarrow a - bx$	$a = 8, b = -4$
	Type 4 (Exp. 2)	$(8 + x) / x \rightarrow a/x + b$	$a = 8, b = 1$

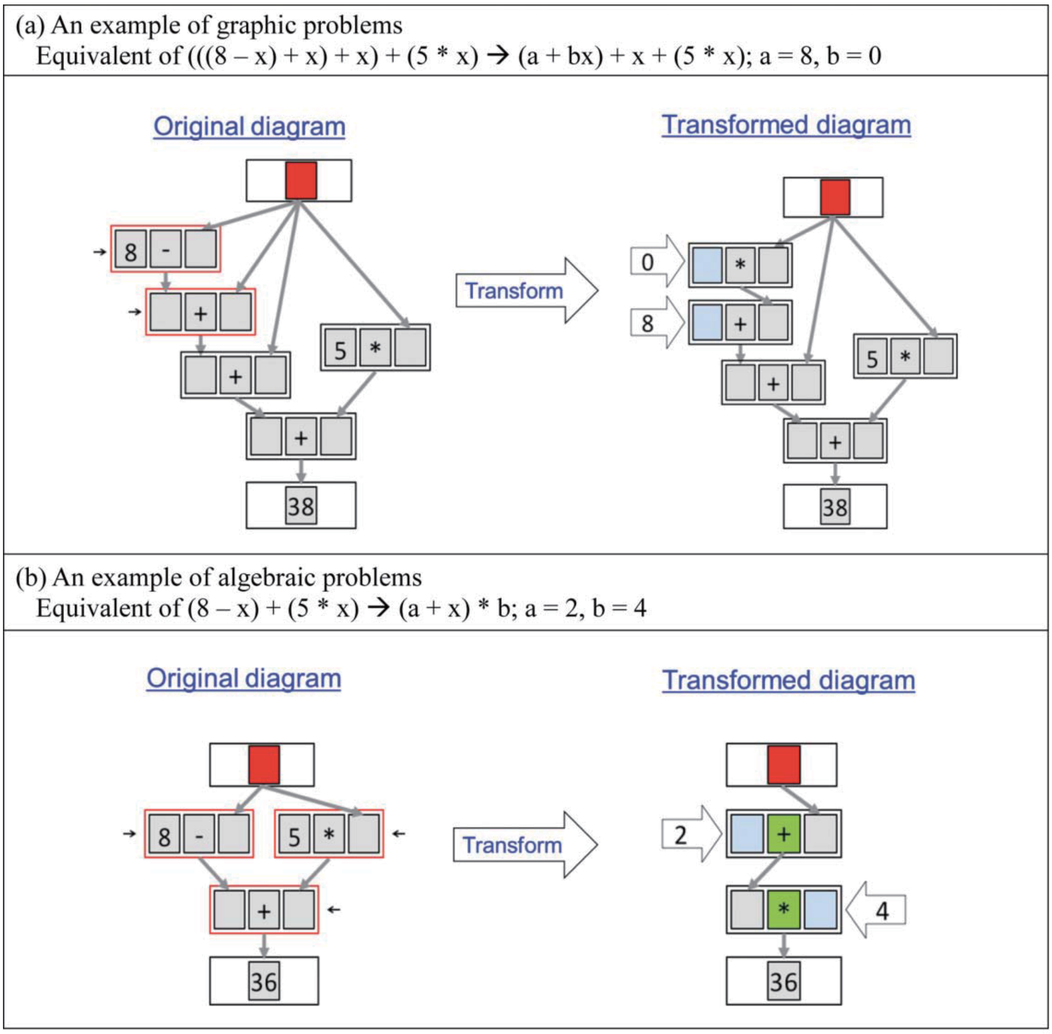


Figure 4. An example of (a) graphic and (b) algebraic problems used in the transfer phase. The left shows the original diagram before transformation, and the right shows the transformed diagram. The correct answers are given in the arrow in this example. (a) is equivalent to transforming $(((8 - x) + x) + x) + (5^*x) \rightarrow (a + bx) + x + (5^*x)$, where $a = 8$, $b = 0$, and (b) is equivalent to transforming $(8 - x) + (5^*x) \rightarrow (a + x)^*b$ where $a = 2$, $b = 4$.

in the left graph. In contrast, the algebraic problems cannot be solved without understanding the equivalence of the two graphs. Thus, graphic problems were a test of how well participants understood how to combine the numbers on the left, while algebraic problems were a test of their understanding of the equivalence of the two structures. Each of graphic and algebraic problems had four subtypes and eight problems were constructed for each subtype, resulting in a total of 64 problems.

Procedure

The study consisted of two phases, a learning phase and a transfer phase. Each phase lasted 2 hours and there were 1 or 2 days between the two phases. The learning phase consisted of two problem sections, one with 20 propagate problems (Figure 1) and the other with 40 simple linearize problems (Figure 2). In the propagate problems, participants learned to propagate numbers up or down. Propagate problems served to familiarize participants with the semantics of the data-flow diagrams and the instructional interface. Also, for the propagate problems, only verbal instruction manipulation could be applied; that is, participants either received explanatory or nonexplanatory verbal instruction when they requested a hint and there was no structural or procedural example provided.

In the learning phase, participants were given a step-by-step example that automatically showed a hint for every step. This example was provided with hints (i.e., both an instructional example and verbal instruction) that corresponded to experimental conditions. Participants could watch how problems were solved in a step-by-step manner. The first problem automatically provided hints for each solution step, but participants had to take actions (i.e., enter numbers, choose boxes) as directed by the system. For the rest of the problems, participants solved problems on their own and experimental manipulations appeared via hints on their request or on errors made.

Across all conditions, an error message immediately appeared whenever participants entered a wrong number (e.g., "6 is not the right answer."). Hints that corresponded to the experimental condition were available on request. If a participant failed to perform a correct step of problem solving for 1 minute, a hint automatically appeared for the corresponding step the participant was performing. After another 1 minute was reached, that step was automatically solved by the system.

To increase motivation participants could earn money for each correct solution step both in learning and transfer phases. The reward amount varied from 1 to 8 cents per step depending on the difficulty of the step. To prevent participants from simply using hints without trying to solve problems,[2] whenever participants asked for hints, 4 cents were deducted. Also, when an error was made, 2 cents were deducted. The reward for a problem would never go below 0. Until all parts were solved correctly, participants could not move on to the next problem.

Transfer phase was conducted only for participants who reached prespecified performance criterion. Because the transfer problems required adjusting procedures already learned from the learning phase, it was impossible to solve the transfer problems without mastering solution procedures during the learning phase. Therefore, only those who made less than an average of 2.5 errors per problem were tested in the transfer phase. This selection criterion was chosen such that we could remove participants who did not figure out the solution rule while still including participants who figured out the rule but made a few computation errors. The criterion of 2.5 or fewer errors allowed us to include the participants who made a couple of computation errors that could be possibly one error for coefficient and one error for constant terms. The transfer phase was identical across all experimental conditions and experimental manipulations occurred only in the learning phase. For transfer task, participants were given 1 minute per problem and their response was followed by a feedback page. Differently from the learning phase, error messages did not appear and hints were not given. Instead, after filling in numbers on the diagram, participants clicked a "done" button and this led to a feedback page for the problem. The feedback page showed the participant's own answers and correct answers simultaneously so that they could compare them with a signal of correct/incorrect. The feedback was presented for 2 seconds for correct and 10 seconds for incorrect responses.

Results and discussion

Learning data

In the learning phase, all participants were successful at completing all 20 propagate problems. Participants solved about 90.05% ($SD = 7.49$) of problems without a single error and made around only 0.13 ($SD = 0.14$) errors per problem. Almost all errors were simple miscalculations. There was no effect of whether they received explanations ($F < 1$) on mean percentage of correctly solved problems. Thus, when the structure of the problem was transparent as in propagate problems, participants were able to master them without instruction.

In contrast, when problems required linearization (transformation), participants showed difficulty learning. There were individual differences in terms of the number of problems solved during the learning phase. In particular, a number of participants had serious problems in the nonexplanatory condition and could not finish entire linearize problems within the 2-hour learning phase. A 2 × 2, between-subjects analysis of variance (ANOVA) was performed to determine the effect of example type and instruction type on the number of linearize problems solved by participants. There was no effect of example type, nor a two-way interaction effect, $Fs < 1$. In contrast, there was a significant effect of instruction type, $F(1, 90) = 20.72$, $p < .001$, $\eta_p^2 = .187$, such that the explanatory group solved significantly more problems than the nonexplanatory group of participants. The explanatory group solved about 8 more problems ($M = 39.39$, $SD = 2.59$) than the nonexplanatory group ($M = 31.36$, *SD* = 11.30) out of 40 linearize problems. This difference was at least in part due to different hint mechanisms between these two groups. In the explanatory conditions, verbal instruction provided final answers of the linearization step if requested by the participant, and thus participants were able to move forward even when they did not really understand problem solutions. In contrast, nonexplanatory participants had to find answers by trial and error because verbal instruction did not provide final answers unless they were timed out.

To control for the number of solved problems and identify initial differences among the conditions, we decided to compare performance on the first 16 linearize problems, but 7 participants in the nonexplanatory condition failed to complete even this many problems (3 structural/nonexplanatory, 4 procedural/nonexplanatory). To remove possible selection bias, 3 participants who showed the greatest number of errors were removed from the explanatory conditions as well. Therefore, the subsequent data analysis included data from 81 participants in total (20 structural/explanatory, 22 structural/nonexplanatory, 18 procedural/explanatory, and 21 procedural/nonexplanatory).

Jarque-Bera tests showed that some of our learning and transfer data were not normally distributed, thus we used arcsine transformation of the data that resulted in normal distribution of our entire data set. Accordingly, we used the transformed data for the subsequent ANOVAs we report. A 2 × 2, between-subjects ANOVA was performed to determine the effect of instructional example and verbal instruction on learning as measured by the correctly performed transformation. Figure 5 shows mean percentages of correctly solved problems without linearization errors out of the first 16 linearize problems during the learning phase. There was no reliable mean difference between the structural and procedural example conditions, $F(1, 77) = 2.06$, $p = .155$, $\eta_p^2 = .026$, implying that both conditions achieved a similar level of learning. In contrast, there was a significant main effect of instruction type in this task, $F(1, 77) = 22.50$, $p < .001$, $\eta_p^2 = .226$. Regardless of example type, participants who were given explanatory verbal instruction ($M = 68.75$, $SD = 15.03$) solved around 25% more problems than the nonexplanatory condition ($M = 43.60$, $SD = 27.80$). There was no instruction-by-example interaction effect, $F < 1$.

Transfer data

Some participants were so confused during the learning phase that they could not deal with the transfer phase. If participants made more than the mean number of 2.5 transformation errors per problem, they did not participate in the transfer phase. This selection criterion removed about 40% of participants from the nonexplanatory conditions. This could have left only high-performing learners in these nonexplanatory conditions (structural/nonexplanatory and procedural/nonexplanatory), whereas all learners were

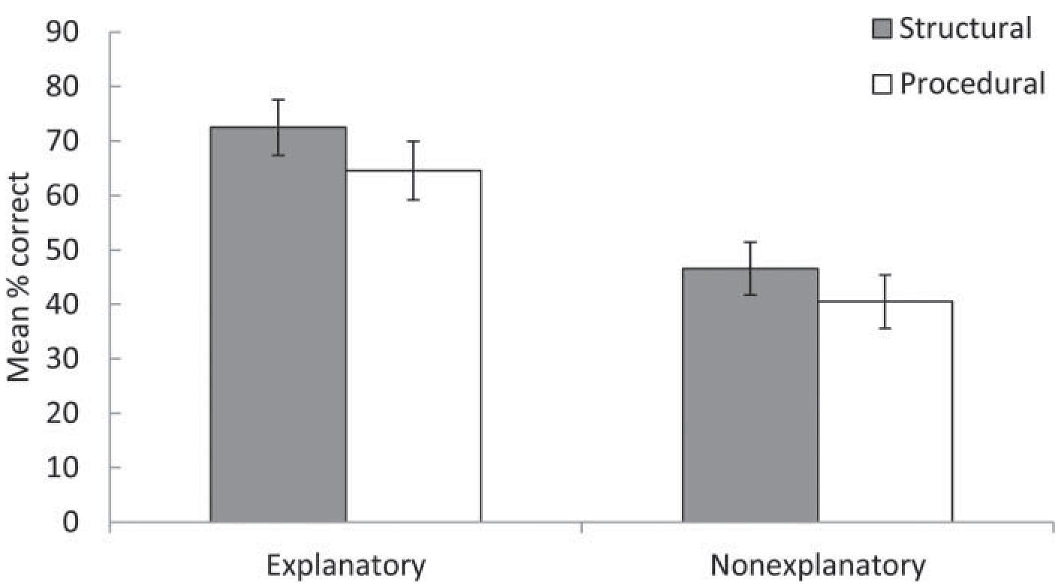


Figure 5. Mean percentages of correctly solved problems in the learning phase of Experiment 1. Error bars represent 1 standard error of mean.

chosen in the other two conditions. To avoid possible selection bias, only top 60% of the participants were chosen in the explanatory conditions as well for transfer data analysis. This finally left 53 participants in total (13 structural/explanatory, 14 structural/nonexplanatory, 12 procedural/explanatory, and 14 procedural/nonexplanatory).[3] In order to see if there was any learning performance difference among these participants, a 2 × 2, between-subjects ANOVA was performed on their last 8 problems of the learning phase. There were not overall effect of example type, $F(1, 49) = 1.68$, $p = .201$, $\eta_p^2 = .033$, effect of instruction type, $F(1, 49) = 0.04$, $p = .835$, $\eta_p^2 = .001$, nor interaction effect between these two factors, $F(1, 49) = 0.16$, $p = .696$, $\eta_p^2 = .003$. Thus, the participants who were analyzed on transfer phase seemed to have reached a similar level of mastery by the end of the learning phase. Participants solved about 72% ($SD = 20.01$) of problems correctly on their last 8 problems of learning phase.

For the transfer phase, problem type (graphic vs. algebraic) was included as a within-subjects variable. A 2 × 2 × 2, mixed ANOVA was performed on transfer performance as measured by the number of correctly solved problems. Different from the results from the learning phase, there was no main effect of verbal instruction, $F(1, 49) = 1.09$, $p = .302$, $\eta_p^2 = .022$, but significant main effect of instructional example, $F(1, 49) = 4.84$, $p = .033$, $\eta_p^2 = .090$. There was not an interaction effect between these two factors, $F(1, 49) = 1.60$, $p = .212$, $\eta_p^2 = .032$. However, there was a significant main effect of problem type, $F(1, 49) = 36.29$, $p < .001$, $\eta_p^2 = .425$. Figure 6 shows mean percentages of correctly solved problems in the transfer phase. Participants overall performed better on graphic problems ($M = 70.46$, $SD = 19.13$) than algebraic problems ($M = 54.83$, $SD = 26.23$). More interestingly, the problem type interacted with the type of instructional example provided in the learning phase, $F(1, 49) = 13.83$, $p = .001$, $\eta_p^2 = .220$. For graphic problems, there was no mean difference between the two example groups, but for algebraic problems, the structural group ($M = 65.16$, $SD = 22.73$) solved more problems correctly than the procedural group ($M = 44.11$, $SD = 25.67$). The problem type did not interact with verbal instructions, $F(1, 49) = 2.95$, $p = .09$, $\eta_p^2 = .057$. There was no three-way interaction, $F(1, 49) = 0.18$, $p = .674$, $\eta_p^2 = .004$.

Discussion

To summarize the results, the instruction type and the example type seemed to have an effect on different aspects of learning. First, the instruction effect was found only when the problem structure was not transparent. Although provision of verbal instruction helped learning for linearize problems, the effect was not observed for propagate problems. In the propagate phase, problem structure appeared to be transparent enough to be understood without verbal instruction. Indeed, many participants immediately knew how to solve propagate problems from the beginning (see also Lee et al., 2011).

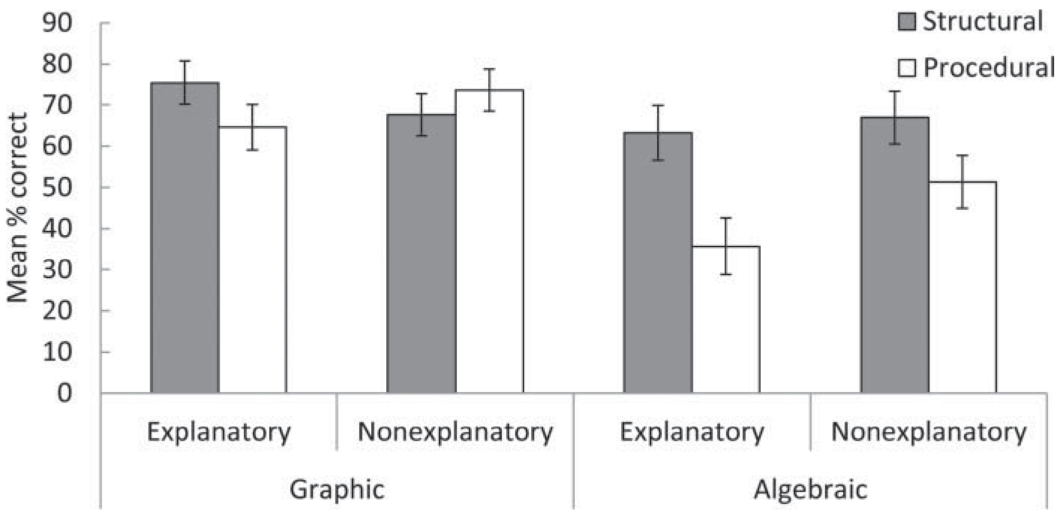


Figure 6. Mean percentages of correctly solved problems in the transfer phase of Experiment 1. Error bars represent 1 standard error of mean.

In contrast, when problem structure was not obvious as in our linearize phase, provision of verbal instruction showed a positive effect by influencing the initial period of learning. Participants who were given explanatory instructions appeared to show less floundering (i.e., making fewer errors) than those who were left to learn just from the examples in the initial period of learning. With verbal instruction, participants solved more problems and made fewer errors initially. Participants who were not given explanatory instructions had to solve problems using trial and error and this seemed to increase incorrect solution searches (i.e., more floundering). However, those participants who could still learn without verbal instruction showed no difference in the later period of learning (i.e., when we analyzed the last 8 linearize problems, there was no difference) and in transfer compared to the best of the participants who were given verbal instruction.

Example type affected transfer performance in that participants who were provided with structural examples transferred learning better to novel problems than those who were provided with procedural examples. Although this manipulation did not seem to have an effect during the learning phase, the example type manipulation appeared to affect understanding of problem structure. When problems required only computational fluency, as in graphic problems, both groups of participants showed comparable transfer performance. In contrast, when problems required understanding the structure, as in algebraic problems, participants who learned with structural examples performed better. These participants better understood how the transformed diagram was equivalent to the original diagram and were more successful in modifying their solution procedure. In contrast, participants who learned with procedural examples seemed to focus more on how to get correct values without an understanding of deep problem structure. As a consequence they were less successful in modifying their procedures for the algebraic problems. This is consistent with the observation that students often perform correct solution steps but show poor transfer performance because of lack of conceptual understanding (Hiebert & Lefevre, 1986; Skemp, 1976).

The reports provided by participants in the debriefing session support this interpretation. Some participants from the procedural example conditions (procedural/explanatory and procedural/nonexplanatory) reported that they were certain that they applied the learned procedure correctly and their answers were correct but the system showed an error message on the feedback page during the transfer phase. These participants often did not notice that problem structure was changed in the algebraic transfer problems, though changed parts were highlighted with different colors. This suggests that those who were provided with procedural examples understood neither the critical structure of the problems nor rationale underlying the solution procedures that they were using. On the other hand, many participants from the structural example conditions (structural/explanatory and structural/nonexplanatory) reported that the color highlighter was very helpful for noticing changed parts of the diagram and adjusting their solution procedure, suggesting that these participants appreciated the key structure of the problem. Indeed, some of these participants mentioned that it was fun to solve these types of problems after they realized that the task was similar to algebra problems. In sum, the opportunity to study problem structure during the learning phase seemed to result in a more flexible use of learned procedures in the transfer phase.

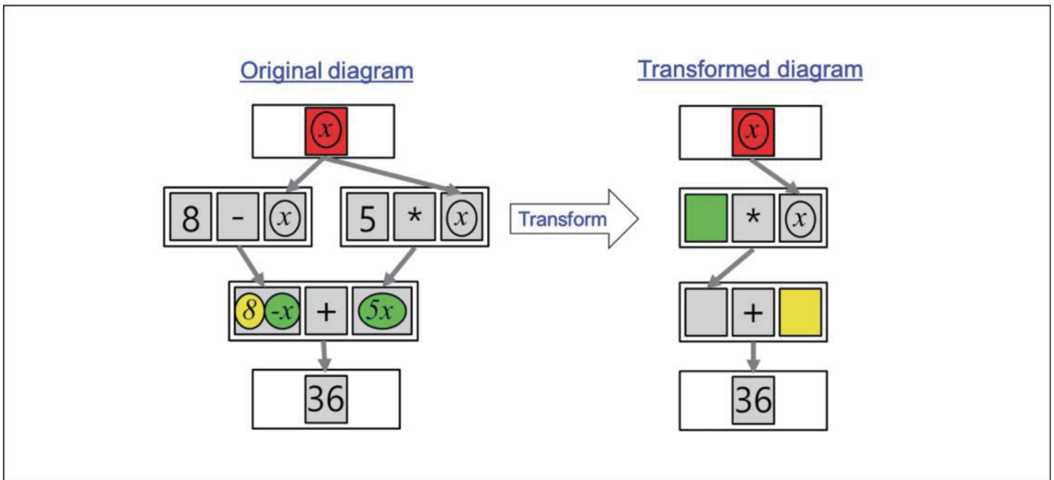
Experiment 2

Based on reports from the first experiment, we thought we could improve our presentation of the examples to make the verbal instruction less critical. Several participants reported that although they understood how the numbers (either in the structural or the procedural example) were propagated in the original diagrams, they did not understand how these numbers related to the numbers they were supposed to enter into the transformed (linearized) diagram. To help resolve this problem, Experiment 2 introduced a special color coding as a visual cue to facilitate a mapping between the diagrams and to help rule inference. Numerous prior studies have emphasized the importance of visual representation in problem solving. For example, Butcher and Aleven (2013) demonstrated that diagram highlighting generated by students supported rule-diagram mapping and led to a better understanding of geometry rules. Jeung, Chandler, and Sweller (1997) also reported that the addition of a visual indicator (flashing highlight) enhanced learning by directing a learner's attention to the relevant part of a diagram. The provision of external visual representations is also known to facilitate scientific problem solving (Mayer, 1989; Mayer & Anderson, 1991, 1992; Lehrer & Schauble, 1998; Penner, Giles, Lehrer, & Schauble, 1996).

Figure 7 shows an example of the color coding used in Experiment 2. Instead of using one single color (gray) for all empty tiles on the linearized diagram, the tile for the coefficient term (the tile next to the operator *) and the tile for the constant term (the tile next to the operator +) were colored differently to correspond to the numbers in the original diagram. We expected that this color coding would help relate the diagrams, especially under the nonexplanatory conditions.

Also, Experiment 2 examined different ways of conveying algebraic interpretation of these diagrams—either as part of the verbal explanation (i.e., algebraic expressions are inserted into verbal instruction) or in the examples. We expected that the type of information should not be important as long as both types of information performed the same function, showing the relation between data-flow diagrams and their algebraic interpretation. Alternatively, we could use both types of information simultaneously. The effect of simultaneous use of the two methods can be either positive or negative. Provision of two sources of information may help students' learning by offering an opportunity to use multiple sources of the information. By integrating multiple sources of information, students can enrich their understanding. On the other hand, multiple sources of information can be simply redundant. Redundancy may cause a split-attention effect and may impair learning (Ward & Sweller, 1990).

In addition to the changes summarized above, we reduced the number of problems in both the learning and the transfer phase. In the learning phase, we reduced linearize problems from 40 to 24. In the transfer phase, we included only algebraic problems in the transfer task because we did not find any effects in the graphic problems. These changes allowed us to control for the number of solved problems across individuals by having all participants finish all the problems within a limited time frame. One consequence of the changes that we introduced into this experiment was that participants



Explanatory verbal instruction: "Imagine the top red box as x and propagate x and numbers down the left and right branches. The sum of the x's goes to the tile next to *, and the sum of the constant numbers goes to the tile next to +. The computations performed by the original (left) diagram and the transformed (right) diagram are the same. That is (8 - x) + 5x = -x + 5x + 8 = 4x + 8. The answer to the green tile is 4 because the sum of the x's is 4 (-x + 5x = 4x). The answer to the yellow tile is 8 because the sum of the numbers is 8 (8 + 0 = 8)."

Nonexplanatory verbal instruction: "Move the mouse to the green tile and enter a number. Move the mouse to the yellow tile and enter a number."

Figure 7. An example of nonverbal scaffolding with expression bubbles. Algebraic expressions were drawn directly on top of the diagram. The coefficient terms were green colored and constant terms were yellow colored for the computation bubbles of the original diagram. The matching colors were used for the corresponding tiles of the transformed diagram.

made more rapid progress through the material, thus they could complete the learning and transfer phases in a single 2-hour experiment, rather than in 2 days as with Experiment 1.

We conducted this study with both college students and persons recruited from an online labor market to see whether our findings could be generalized to other populations. Several studies have reported that an instructional method that is effective with experienced learners might not be effective with inexperienced learners (e.g., Kalyuga, 2007, Kalyuga et al., 2003; Salden et al., 2010). One may predict that providing a combination of verbal instruction and nonverbal scaffolding may help low-ability learners by providing multiple sources of information that they can study from. In contrast, it may harm high-ability learners by causing a redundancy effect (Mayer, Heiser, & Lonn, 2001; Moreno & Mayer, 2002) or a split-attention effect (Tarmizi & Sweller, 1988; Mwangi & Sweller, 1998; Ward & Sweller, 1990). By testing two different populations we will see how different types of instruction impacts success in learning and how optimal instruction design can change. In the following method and result sections, we will report findings from both populations.

Method

Participants

Two hundred and forty-three subjects participated in the study. We collected data from CMU ($N = 80$) and from an online labor market, Amazon Mechanical Turk ($N = 163$). Out of 243 participants, 81 wanted to quit the study at various phases of the study. Table 2 shows the number of participants who dropped out in each phase of the study for each experimental condition. Most withdrawals occurred among the Mechanical Turk participants (77 out of 81 cases). Many of those (41) decided to quit the study while solving propagate problems and others wanted to quit in the linearize section (34) or transfer phase (2). We constrained our analysis to those who made it past propagate session because experimental manipulations occurred only after this session. From CMU, the participant pool consisted of graduate and undergraduate students (35 male and 45 female, $M = 22.08$ years, $SD = 2.76$). Seventy-two of the 80 participants reported their SAT math scores ($M = 737.24$). Participants received $10/hour plus a performance-based bonus. From Mechanical Turk, 122 participants made it past propagate problems (72 male and 50 female, $M = 27.14$ years, $SD = 6.28$) and they reported various levels of education background (47% 4-year college degree, 31% 2-year degree or some college, 16% master's degree, 4% high school diploma, and 2% professional degree or certification). Among the Mechanical Turk population, only 54 out of 122 subjects reported their SAT math scores ($M = 644.97$). Online participants received the fixed amount of $5 plus a performance-based bonus. Each participant was randomly assigned to one of the four experimental conditions.

Table 2. Number of participants who quit the study in various phases of the experiment and those who completed the study in Experiment 2. Because there was no experimental manipulation in the propagate session, the reported numbers in the propagate column represent the number of withdrawals regardless of experimental conditions.

		Drop-out					
		Propagate	Linearize	Transfer	Completed	Removed	Final sample
Carnegie Mellon University	explanatory/scaffolded	0	0	0	20	2	18
	explanatory/nonscaffolded		0	0	20	2	18
	non-explanatory/scaffolded		2	0	18	0	18
	nonexplanatory/nonscaffolded		2	0	18	0	18
Mechanical Turk	explanatory/scaffolded	41	6	1	21	3	18
	explanatory/nonscaffolded		5	0	22	5	17
	nonexplanatory/scaffolded		12	1	22	0	22
	nonexplanatory/nonscaffolded		11	0	21	1	20
Total		41	38	2	162	13	149

Design and materials

A 2 × 2, between-subjects design was employed. The first independent variable was the type that included verbal instruction (explanatory vs. nonexplanatory). In the explanatory condition, detailed verbal instruction was provided with algebraic expressions to explain the structure of the diagram and how to determine values to fill in on the linearized diagram. In the nonexplanatory condition, information was provided on only general interface issues in the step of the linearize task. An example of text hints given for each condition is shown in Figure 7.

Crossed with the manipulation of verbal instruction, the presence of nonverbal scaffolding was also manipulated (scaffolded vs. nonscaffolded). For the scaffolded condition we used the structural example condition of the previous experiment but now with color coding like the ones shown in Figure 7; for example, in Figure 7 (left), all propagated numbers rejoin in the rectangular box that has (8 − x) and 5x. Here, variable terms −x and 5x were green while the constant term 8 was yellow. These colors were matched with the tile color of the linearized diagram (see Figure 7 (right)); that is, the tile for the coefficient term was green (same as the color of -x and the 5x bubble), and the tile for the constant term was yellow (same as the color of 8). This color coding was intended to help participants understand where exactly the answers were derived from the expressions. In the nonscaffolded condition, there was no example provided.

For the transfer task, different from Experiment 1, all transfer problems were algebraic subtype of problems. We observed that some participants figured out the pattern of the answers (i.e., simply changing the sign of the constant number) for one type of algebraic process based on the feedback after a few trials, without necessarily understanding why they had to make that adjustment on their answers. Therefore, a new type of problem was constructed and it is shown in Table 1. There were four subtypes, and eight problems were constructed for each subtype, resulting in a total of 32 problems.

Procedure

The study consisted of two phases, a learning phase and a transfer phase. Participants solved 20 propagate problems and 24 linearize problems in the learning phase. Because we did not find any instructional effects on propagate problems in Experiment 1, we did not apply any experimental manipulations to the propagate problems. All participants received the same kinds of hint texts as in the explanatory condition of Experiment 1. The 2 × 2 experimental manipulations were administered only for the linearization part of the task during the learning phase. Immediately after the learning phase, participants were tested with transfer problems. All other procedures were identical to those of Experiment 1. The entire experiment took about 2 hours.

Results and discussion

Learning data

As shown in Table 2, we had different numbers of withdrawals across the two populations and the various experimental conditions. A chi-square test revealed a significant relation between the withdrawal rate and population, $X^2(1, N = 202) = 16.55$, $p < .001$, such that significantly more withdrawals occurred among the Mechanical Turk pool than the CMU pool. Also, we found that the nonexplanatory group was significantly more likely to quit the study than the explanatory group in the middle of the linearize section, $X^2(1, N = 202) = 6.14$, $p = .013$. This pattern was observed in both the CMU group (0 explanatory vs. 4 nonexplanatory drop-outs) and the Mechanical Turk group (11 explanatory vs. 23 nonexplanatory drop-outs). As in Experiment 1, this was perhaps due to different hint mechanisms between the explanatory and the nonexplanatory conditions. Only explanatory participants were able to get answers of problems on their hint request and this enabled them to go forward even when they were not able to solve the problem on their own. Different numbers of withdrawals from each condition could have left only high-performing participants in some conditions whereas all participants were included in some others. As in Experiment 1, to remove possible selection bias, some participants were removed from some conditions. Among the CMU population, four participants who showed the worst learning performance (2 explanatory/scaffolded, 2 explanatory/nonscaffolded) were removed from the explanatory conditions so that only the top 90% of participants were chosen in each experimental condition. Among the Mechanical Turk population, nine more subjects (3 explanatory/scaffolded, 5 explanatory/nonscaffolded, and one nonexplanatory/nonscaffolded) were removed so that only the top 60% of participants were chosen in each experimental condition. Therefore, the subsequent data analysis included data from 149 participants in total. Table 2 shows the number of participants who were included in the final data analysis. The final sample consisted of 72 CMU students (32 male and 40 female, $M_{\text{age}} = 22$, $M_{\text{SAT(N = 64)}} = 767.66$) and 77 Mechanical Turk participants (45 male and 32 female, $M_{\text{age}} = 26.62$, $M_{\text{SAT(N = 33)}} = 646.49$; 47% with a 4-year degree, 29% with a 2-year degree or some college, 17% with a master's degree, 5% high school graduates, and 2% with a professional degree or certification).

Jarque-Bera tests showed that some of our learning and transfer data were not normally distributed, thus we used arcsine transformed data that resulted in a normal distribution for the subsequent ANOVAs we report. Regarding the performance on propagate problems, a 2 × 2 × 2, between-subjects ANOVA was performed. Three between-subjects variables were population (CMU vs. Mechanical Turk), verbal instruction (explanatory vs. nonexplanatory), and nonverbal scaffolding (scaffolded vs. nonscaffolded). Because there were no instructional manipulations on propagate problems, we did not expect any performance difference between experimental conditions. As expected, there were not main effects of either instruction type, $F(1, 141) = 1.54$, $p = .217$, $\eta_p^2 = .011$, nor nonverbal scaffolding, $F < 1$. Also, the performance of one population was not different from that of the other, $F(1, 141) = 1.19$, $p = .277$, $\eta_p^2 = .008$. There were no two-way or three-way interaction effects, Fs < 1. Overall, participants solved about 87% ($SD = 9.46$) of propagate problems correctly without an error.

Regarding the performance on linearize problems, we analyzed the number of problems where participants correctly performed transformation without making errors or requesting hints. Figure 8 shows mean percentages of problems solved with a correct linearize step out of the 24 linearize problems in the learning phase. The CMU participants ($M = 62.38$, $SD = 27.64$) correctly performed linearization steps more often than the Mechanical Turk participants ($M = 43.24$, $SD = 32.26$), $F(1, 141) = 18.87$, $p < .001$, $\eta_p^2 = .118$. There were significant main effects of verbal instruction, $F(1, 141) = 58.62$, $p < .001$, $\eta_p^2 = .294$, and of nonverbal scaffolding, $F(1, 141) = 12.16$, $p = .001$, $\eta_p^2 = .079$. More interestingly, there was a significant interaction effect of verbal instruction by nonverbal scaffolding manipulations, $F(1, 141) = 14.56$, $p < .001$, $\eta_p^2 = .094$. This interaction was due to significantly worse performance in the nonexplanatory/nonscaffolded condition ($M = 22.70$, $SD = 26.09$) than in the other three conditions (averaged

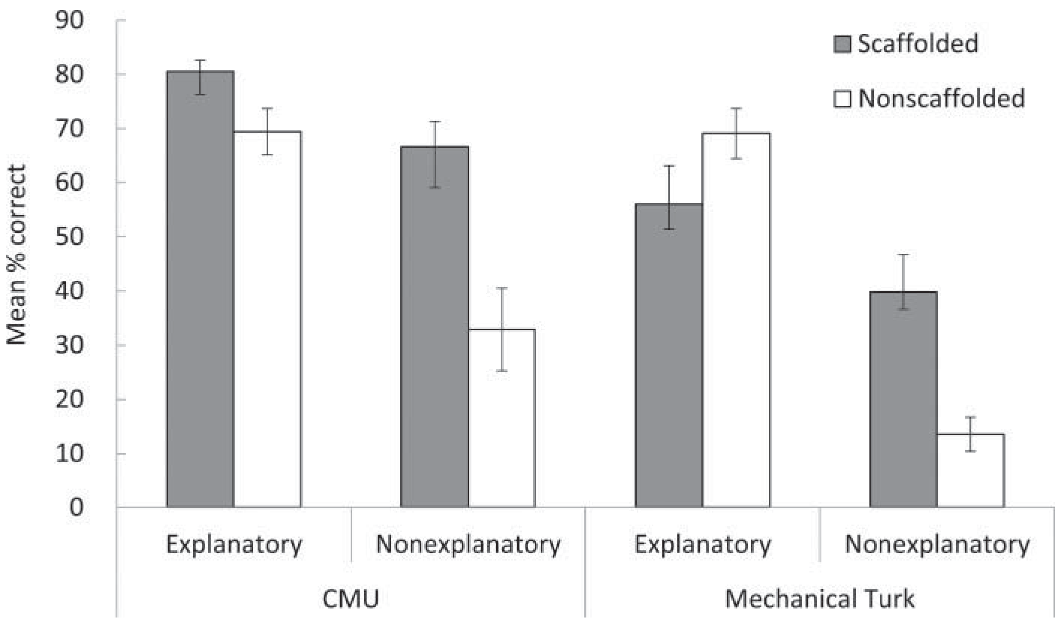


Figure 8. Mean percentages of problems solved without linearization errors out of 24 problems in the learning phase of Experiment 2. Error bars represent 1 standard error of the mean.

$M = 62.69$, $SD = 26.37$). Also, the population variable interacted with nonverbal scaffolding, $F(1, 141) = 4.21$, $p = .042$, $\eta_p^2 = .029$. This seemed to be because the nonverbal scaffolding effect was bigger among the university population than the Mechanical Turk population. In the CMU population, nonverbal scaffolding had about a 22% performance increase whereas in the Mechanical Turk population that received nonverbal scaffolding had only about a 8% of increase. There was no interaction between population and instruction type, $F(1, 141) = 2.13$, $p = .147$, $\eta_p^2 = .015$, nor a three-way interaction effect, $F(1, 141) = 0.73$, $p = .395$, $\eta_p^2 = .005$.

We were also interested in whether verbal instruction or nonverbal scaffolding would be more effective for learning. Different patterns of results were observed between the two subgroups of population, $F(1, 71) = 6.35$, $p = .014$, $\eta_p^2 = .082$. Among the CMU participants, there was virtually no difference between the explanatory/nonscaffolded condition ($M = 69.44$, $SD = 18.08$) and the nonexplanatory/scaffolded condition ($M = 66.67$, $SD = 19.80$), $t(34) = 0.46$, $p = .652$. Either verbal instruction or nonverbal scaffolding seemed to successfully communicate the underlying structure of these problems. However, among the Mechanical Turk participants, those under the explanatory/nonscaffolded condition ($M = 69.12$, $SD = 18.98$) solved significantly more problems correctly than those under the nonexplanatory/scaffolded condition, ($M = 39.77$, $SD = 32.37$). For this subgroup of population, verbal instruction was more effective than nonverbal scaffolding for learning, $t(37) = 3.30$, $p = .002$.

It is also interesting to see whether the provision of multiple sources of information is more effective than the provision of a single source of information. Again, different patterns of results were found between the two subgroups of population, $F(2, 105) = 3.71$, $p = .028$, $\eta_p^2 = .066$. The CMU participants solved significantly more problems correctly when both verbal instruction and scaffolding were provided than when there was only one type of information provided. Explanatory/scaffolded participants ($M = 80.56$, $SD = 8.81$) performed better than either explanatory/nonscaffolded, $t(34) = 2.32$, $p = .027$, or nonexplanatory/scaffolded participants, $t(34) = 2.65$, $p = .012$. However, performance differences were not found among the Mechanical Turk participants between multiple and single source of information groups, $F < 1$.

Transfer data

Because there were not any significant main effects or interaction effects regarding solution time, our data analyses focused on the number of correctly solved problems. Figure 9 shows the mean percentages of correctly solved problems out of 32 transfer problems. In general, similar patterns of results were obtained as in the learning phase. The CMU participants ($M = 42.40$, $SD = 27.64$) solved significantly

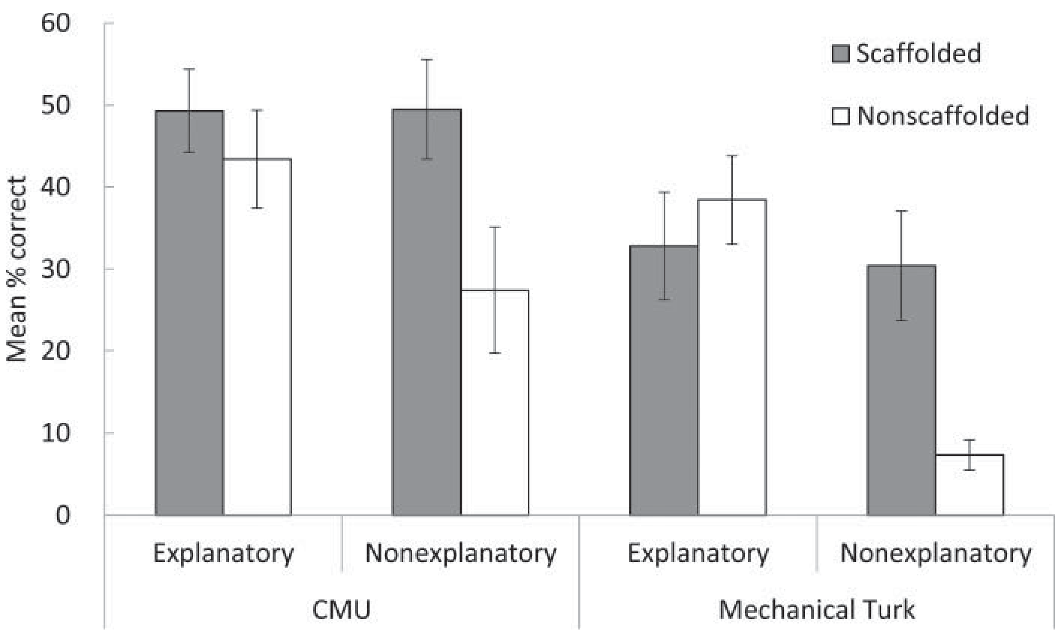


Figure 9. Mean percentages of correctly solved problems out of 32 problems in the transfer phase of Experiment 2. Error bars represent 1 standard error of the mean.

more transfer problems correctly than the Mechanical Turk participants ($M = 26.75$, $SD = 26.54$), $F(1, 141) = 10.75$, $p = .001$, $\eta_p^2 = .071$. There was a significant main effect of verbal instruction, $F(1, 141) = 11.76$, $p = .001$, $\eta_p^2 = .077$, and nonverbal scaffolding, $F(1, 141) = 7.12$, $p = .009$, $\eta_p^2 = .048$. More interestingly, there was a significant interaction effect of verbal instruction by nonverbal scaffolding manipulations, $F(1, 141) = 7.96$, $p = .005$, $\eta_p^2 = .053$. As in the learning phase, this interaction was due to significantly worse performance in the nonexplanatory/nonscaffolded condition ($M = 16.86$, $SD = 25.08$) than the other three conditions (averaged $M = 40.29$, $SD = 26.64$). The other three conditions did not differ from each other, $F < 1$. There was no other two-way interaction nor a three-way interaction effect, Fs < 1, implying that patterns of results were identical among the different subgroups of population. Therefore, in the transfer phase, different from the learning phase, we did not find any partial benefits of one type of information over the other nor benefits of providing multiple sources of information over providing a single source of information in either subgroup of the two populations.

Discussion

To summarize, we found somewhat different patterns of results between the two subgroups of population in the learning phase and very similar patterns of results in the transfer phase. First of all, the CMU participants performed better than the Mechanical Turk participants in both learning and transfer phases. In the learning phase, although both subgroups showed the biggest learning deficit in the nonexplanatory/nonscaffolded condition, relative benefits of verbal instruction and nonverbal scaffolding seemed to be different between the two populations. In this experiment, we were interested in whether verbal instruction or nonverbal scaffolding would be more effective for learning. Verbal instruction without nonverbal scaffolding could be at a disadvantage in that participants have to mentally map verbally provided expressions onto the diagrams, although the explanations do explicitly state how to compute each term for the linearization task. In contrast, nonverbal scaffolding without explanations could be at a disadvantage in that the rules are not explicitly stated, although the mapping between algebraic expressions and diagrams is directly given. The result patterns were different between the two populations. Among the CMU participants, nonverbal scaffolding was as effective as verbal instruction. However, among the Mechanical Turk participants, verbal instruction was more effective than nonverbal scaffolding. The CMU participants appeared to be able to interpret nonverbal scaffolding, whereas Mechanical Turk participants seemed to have a harder time interpreting expressions drawn on the diagram.

In addition, we found that provision of verbal instruction helped participants persist through the task and it was especially helpful to the Mechanical Turk participants. As shown in different rates of withdrawal (see Table 2), the nonexplanatory group was significantly more likely to quit the study than the explanatory group regardless of whether nonverbal scaffolding was provided. Such patterns were observed in both populations. Among the CMU participants, withdrawals were observed only in the nonexplanatory conditions, and among the Mechanical Turk participants, nonexplanatory conditions showed twice as many withdrawals as explanatory conditions. When a task was challenging, learners appeared to rely more on verbal instruction that explicitly told learners what to do and so avoid inducing solution procedures from nonverbal scaffolding.

Also, combining verbal instruction and nonverbal scaffolding (explanatory/scaffolded condition) resulted in better performance in the learning phase than provision of one single instructional feature (either explanatory/nonscaffolded or nonexplanatory/scaffolded) only among the CMU participants. Providing both verbal and nonverbal forms of instruction could help learning by allowing learners to have access to multiple sources of information or harm learning by creating split-attention effect (Mwangi & Sweller, 1998; Tarmizi & Sweller, 1988; Ward & Sweller, 1990). If Mechanical Turk participants had a hard time interpreting nonverbal expressions drawn on the diagram, it would be better for these learners to ignore the nonverbal scaffolding and focus on verbal instruction by putting themselves into an explanatory/nonscaffolded condition rather than an explanatory/scaffolded condition. Indeed, there was no performance difference between these two groups among the Mechanical Turk participants.

However, all of these learning differences between the two populations disappeared in the transfer phase. Both subgroups of participants showed a deficit only when there was nothing to convey problem structure (i.e., nonexplanatory/nonscaffolded). The most striking outcome of this experiment was the equivalent transfer performance of groups given verbal explanation, nonverbal scaffolding, or both. While the verbal instruction was clearly helpful in the absence of scaffolding, and scaffolding was clearly helpful in the absence of verbal instruction, they basically were redundant. When an algebraic expression was provided, the structure of the problem appeared to become apparent and the verbal instruction seemed unnecessary. The nonverbal scaffolding never provided explicit problem solving rules, but participants were able to find rules for themselves using algebraic expressions and this led to better transfer outcomes. This finding is also interesting in that a symbolic representation was able to facilitate the understanding of a visual representation. Many teachers use visual representation to help students understand mathematical symbolic notations. When learners are capable of symbolic reasoning like the college-level students in our study, symbolic representation also can be used to help understanding of other representations. This is consistent with the idea of the "power of symbols" (Arcavi, 1994). Arcavi describes examples of symbol sense (compared to number sense) and claims that understanding of the power of symbols is part of symbol sense. People with good symbol sense would know when and how symbols are used to represent relations that are otherwise invisible. In our study, making connections to algebraic expressions allowed participants to appreciate the deep structure of the problem that is otherwise hard to make sense of.

General discussion

Through two experimental studies, we examined the effect of embellishing problem-solving examples with problem structure and the effect of providing verbal instruction on learning and transfer in mathematical problem solving. Both experiments showed that delivery of deep problem structure mattered. In Experiment 1, while participants learned equally well with either structural or procedural examples, the transfer results showed there were significant differences in what they had learned and these differences were not impacted by the presence of verbal instructions. Experiment 2 showed poor performance in learning and transfer when there was no instructional support to convey problem structure (via either verbal instruction or scaffolded example).

The research also showed large effects of verbal instruction on learning. In both experiments significantly more participants with verbal instruction made it to the transfer phase. However, among the participants who completed the study, the addition of verbal instruction did not improve transfer performance over just a scaffolded example (and equally addition of a scaffolded example did not improve performance over just verbal instruction). Figures 6 and 9 reveal that this is not a ceiling effect and that there were effects of other variables.

All together, our study supports several instructional principles that could be applied to computer-based learning environments. First, it is critical to convey problem structure in order for students to acquire better problem schema, which in turn results in better transfer performance. As shown, the success or failure of learning conditions heavily depends on how well instructional features communicate problem structure. This is consistent with numerous studies that have shown the importance of emphasizing deep problem structure (e.g., Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1983; Jitendra et al., 2011; Perry, 1991). In addition to confirming previous findings, we examined when verbal instruction helps or does not help learning. While we never found an advantage of deleting verbal instructions, the second experiment suggests that a properly scaffolded example can be sufficient in itself, particularly when it comes to what a participant takes away from a learning phase and reveals in a transfer phase. These results suggest a possible explanation for why we have so many mixed results for providing any kinds of instruction in previous literature. Provision of instructional explanations does not guarantee a positive learning outcome (for reviews, see Wittwer & Renkl, 2008; 2010) and its empirical evidence is mixed (e.g., Atkinson, 2002; Gerjets, Scheiter, & Catrambone, 2006; Große & Renkl, 2006; Lovett, 1992; Renkl, 2002; Ward & Sweller, 1990). According to our study results, an instructional explanation will not help when problem structure is transparent enough for students to understand the problem solution. When the problem structure is not transparent, verbal instruction will help by explicitly guiding students to what they have to do and reducing effort spent on unnecessary search processes. Verbal instruction may also be helpful for those students who cannot understand the problem structure; for instance, those nonexplanatory subjects in our studies who seemed to get lost might not have been able to interpret the scaffolding that we provided.

Second, provision of instruction appeared to play a critical role in having learners persist through a challenging task. In Experiment 2, more participants wanted to quit the study when they were given nonexplanatory instruction than when they were given explanatory instruction; this applied to both subgroups of population. When verbal instructions are given, subjects know that the explanation is there to help them even if they do not immediately understand the instructions, but subjects may not be aware of the significance of nonverbal features. Given that Mechanical Turk participants showed a greater number of drop-outs across the study, it may be worth thinking about why the verbal instruction especially encouraged the Mechanical Turk participants to persist. One possible explanation is that the Mechanical Turk participants were less motivated than CMU participants and verbal instructions may be especially helpful to less motivated learners. In an online setting, there is a greater opportunity to leave, compared to a typical lab setting in which a subject has to interact with an experimenter in person. When students are more or less motivated to learn, their study strategy and effective instructional methods may change. It is well known that there is a powerful link among motivation, learning, and academic achievement (e.g., Deci, Vallerand, Pelletier, & Ryan, 1991; Dweck, 1986). Future studies need to investigate how students' persistence on a task is affected by different types of instructional methods and how effectiveness of verbal instruction interacts with students' motivation.

Third, this research suggests that benefits of verbal instruction versus nonverbal scaffolding might be different depending on different populations of learners. Although we found very similar patterns of results between the two subgroups of population in the transfer phase of Experiment 2, effects of verbal versus nonverbal methods were different in the learning phase. Specifically, verbal instruction was more effective than the nonverbal scaffold to Mechanical Turk participants, whereas the nonverbal scaffold was as effective as the verbal instruction to CMU participants. Our two subgroups of participants may differ in several ways. As we pointed out earlier, they may have had different levels of motivation and such different motivation may cause different levels of cognitive effort in interpreting nonverbal scaffolding features. An alternative explanation is that the difference in relative effectiveness of instructional methods could arise from different mathematical ability. In Experiment 1, provision of answers on hint requests appeared to reduce floundering when participants just started learning, and this suggests that verbal instruction can help learning when learners are not experienced enough for the learning materials. Likewise, in Experiment 2 verbal instruction could be more beneficial to Mechanical Turk participants because they are less experienced than CMU participants. The CMU students presumably have higher math ability than the Mechanical Turk participants given that the latter group reported various levels of educational background and a lower mean of SAT math scores (note that only a small number of Mechanical Turk participants reported their scores). Also, the CMU participants showed better performance in both learning and transfer phases. If the CMU group indeed had higher math ability, then we can suggest that instructional design choices drawn on by higher-ability learners could lead to inferior performance if chosen by lower-ability learners, who are indeed in greater need of instructional support. Although this result is consistent with expertise reversal effect (Kalyuga, 2007; Kalyuga et al., 2003), this is the first study that has shown possibly greater benefits of verbal instruction over nonverbal instruction to inexperienced learners. In addition, performance differences between the two populations offers us a cautionary message that experimental findings found in a lab setting may not be applied to other populations. Instructional design choices are likely to have different implications for different populations of learners.

The challenge of appreciating deep structure and inducing a correct rule intersects with one of the central controversies in education, which is how much instructional guidance needs to be provided in a learning environment (Kirschner, Sweller, & Clark, 2006; Koedinger & Aleven, 2007; Kuhn, 2007; Lee & Anderson, 2013; Mayer, 2004; Tobias & Duffy, 2009). While allowing students to find their own solutions may have advantages, a typical criticism of discovery-oriented approaches is that students flounder trying to find the solution and may never be able to discover what they are to learn. This would be a particular danger when problems have critical structure that is not obvious; this is what happened when Mechanical Turk participants wanted to quit in the middle of the study. The current study suggests that amount of instruction and type of instructional supports need to be adjusted for different learners.

Limitations and future directions

Additional research is needed because of possible limitations of the current study. First, the current study used isomorphs of algebra problems to gain more control over background knowledge so that all participants would start out without knowing how to solve problems. However, this may not be generalized to actual algebra learning beyond a laboratory context. The specific features of instructional supports we adopted (e.g., color coding) were sometimes dependent on unique characteristics of the diagram task. Also, as we have shown that effective instructions can differ depending on populations of learners, our instructional design choices may not work for algebra students. Future work will need to test how general implications of the current study can be changed into applicable guidelines to actual algebra learning materials. One of our preliminary efforts made experimental modifications to the Cognitive Algebra Tutor (Lee et al., 2013) by focusing on how to highlight the critical structure of equation solving problems using a special color coding.

Second, in Experiment 2 we tested different types of instructional methods with a college population and a Mechanical Turk population. Although overall patterns of the results were similar, we identified several performance differences between these two populations. We related the observed performance difference with different levels of motivation due to an offline versus an online setting and math ability due to educational background and reported SAT scores. However, we did not systematically check the motivation level or math ability of these populations. Therefore, future studies will need to investigate how different levels of motivation and/or ability interact with different instructional designs in a more systematic way. Although one may reasonably argue that Mechanical Turk participants are different from CMU participants, careful interpretation will be needed to generalize findings to other populations. In an online labor market, people choose to participate in a study after reading a brief description of the study and this could have caused inclusion of online participants who have a relatively high interest in math. Also, many of Mechanical Turk participants wanted to quit the study in various phases of the study. This could have left high performing participants. (To remedy this problem, we removed low performing participants proportionally from all of the conditions. The results do not change if we do not remove these subjects). Nevertheless, different withdrawal rates from different instructional conditions give us an idea of which type of instructional design is better for having learners persist through a challenging task.

Lastly, more research is needed to investigate factors that make problem structure apparent to learners. Clearly problem features and instructions will play an important role. In the current study, we specifically used learner's prior knowledge of algebra to provide the link for making the problem structure apparent. As long as students had a basic knowledge of algebra, which our participants did, algebraic expressions made a good reference for understanding this deep structure of data-flow diagrams. However, if learners had never been exposed to algebra, then they would not have been able to take advantage of algebra-rich information. For instructions to be helpful, they should be provided in accord with the development of children's understanding (Carpenter & Lehrer, 1999).

Notes

1. All experimental materials are available at http://tinyurl.com/nkpl2br
2. Such a "hint abuse" strategy (Aleven & Koedinger, 2000; Aleven et al., 2004) would only make sense in the instruction condition in which the hints provided information on how to solve the problem.
3. The basic results do not change without this removal of low performers from the explanatory conditions. Also, the difference between explanatory and nonexplanatory groups during learning (Figure 5) remains.

Funding

This project was funded by the Institute of Education Sciences (R305A100109) and by a National Research Foundation of Korea Grant (NRF-2014S1A5A8018026).

References

Arcavi, A. (1994). Symbol sense: Informal sense-making in formal mathematics. *For the Learning of Mathematics*, *14*, 24—35.

Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Proceedings of the fifth international conference on intelligent tutoring systems, ITS 2000* (pp. 292—303). Berlin, Germany: Springer Verlag.

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2004). Toward tutoring help seeking: Applying cognitive modeling to meta-cognitive skills. In J. C. Lester, R. M. Vicario, & F. Paraguaçu (Eds.) *Proceedings of seventh international conference on intelligent tutoring systems, ITS 2004* (pp. 227—239). Berlin, Germany: Springer Verlag.

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, *94*, 416—427.

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.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612—637.

Bisanz, J., & LeFevere, J. A. (1992). Understanding elementary mathematics. In J. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 113—136). Amsterdam, Netherlands: North Holland, Elsevier Science.

Brunstein, A., Betts, S., & Anderson, J. R. (2009). Practice enables successful learning under minimal guidance. *Journal of Educational Psychology*, *101*, 790—802.

Butcher, K. R., & Aleven, V. (2013). Using student interactions to foster rule–diagram mapping during problem solving in an intelligent tutoring system. *Journal of Educational Psychology*, *105*(4), 988–1009.

Carpenter, T. P., & Lehrer, R. (1999). Teaching and learning mathematics with understanding. In E. Fennema & T. R. Romberg (Eds.), *Mathematics classrooms that promote understanding* (pp. 19—32). Mahwah, NJ: Lawrence Erlbaum.

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology*, *86*, 360—367.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161—238). Hillsdale, NJ: Lawrence Erlbaum.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145—182.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121—152.

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist*, *26*, 325—346.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*, 1040—1048.

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D. … Zumeta, R. O. (2009). Remediating number combination and word problem deficits among students with mathematics difficulties: A randomized control trial. *Journal of Educational Psychology*, *101*, 561—576.

Fuchs, L. S., Seethaler, P. M., Powell, S. R., Fuchs, D., Hamlett, C. L., & Fletcher, J. M. (2008). Effects of preventative tutoring on the mathematical problem solving of third-grade students with math and reading difficulties. *Exceptional Children*, *74*, 155—173.

Fuchs, L. S., Zumeta, R. O., Schumacher, R. F., Powell, S. R., Seethaler, P. M., Hamlett, C. L., & Fuchs, D. (2010). The effects of schema-broadening instruction on second graders' word-problem performance and their ability to represent word problems with algebraic equations: A randomized control study. *Elementary School Journal*, *110*, 440—463.

Fuson, K. C., & Willis, G. B. (1989). Second graders' use of schematic drawings in solving addition and subtraction word problems. *Journal of Educational Psychology*, *81*, 514—520.

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

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction*, *16*, 104—211.

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

Große, C. S., & Renkl, A. (2006). Effects of multiple solution methods in mathematics learning. *Learning and Instruction*, *16*, 122—138.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1—27). Hillsdale, NJ: Erlbaum.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory and Cognition*, *15*(4), 332—340.

Jeung, H., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology*, *17*, 329–345.

Jitendra, A. K., Star, J., Dupuis, D. N., & Rodriguez, M. (2013). Effectiveness of schema-based instruction for improving seventh-grade students' proportional reasoning: A randomized experiment. *Journal of Research on Educational Effectiveness*, *6*, 114—136.

Jitendra, A. K., Star, J. R., Rodriguez, M., Lindell, M., & Someki, F. (2011). Improving students' proportional thinking using schema-based instruction. *Learning and Instruction*, *21*, 731—745.

Jitendra, A. K., Star, J., Starosta, K., Leh, J., Sood, S., Caskie, G. … Mack, T. R. (2009). Improving students' learning of ratio and proportion problem solving: The role of schema-based instruction. *Contemporary Educational Psychology*, *34*, 250—264.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*, 509—539.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, *38*, 23—31.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75—86.

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*, 30—43.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*, 239—244.

Kuhn, D. (2007). Is direct instruction an answer to the right question? *Educational Psychologist*, *42*, 109—113.

Lee, H. S., & Anderson, J. R. (2013). Student learning: What has instruction got to do with it? *Annual Review of Psychology*, *64*, 445—469.

Lee, H. S., Anderson, J. R., Berman, S. R., Ferris-Glick, J., Joshi, A., Nixon, T., & Ritter, S. (2013, September). Exploring optimal conditions of instructional guidance in an algebra tutor. Paper presented at the annual meetings of the Society for Research on Educational Effectiveness (SREE). Washington, DC.

Lee, H. S., Anderson, A., Betts, S., & Anderson, J. R. (2011). When does provision of instruction promote learning? In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 3518—3523). Austin, TX: Cognitive Science Society.

Lee, H. S., Betts, S., & Anderson, J. R. (2015). Not taking the easy road: When similarity hurts learning. *Memory and Cognition*, *43*(6), 939—952.

Lee, H. S., Fincham, J., Betts, S., & Anderson, J. R. (2014). An fMRI investigation of mathematical problem solving. *Trends in Neuroscience and Education*, *3*, 50—62.

Lehrer, R., & Schauble, L. (1998). Reasoning about structure and function: Children's conceptions of gears. *Journal of Research in Science Teaching*, *35*, 3–25.

Leinhardt, G. (2001). Instructional explanations: A commonplace for teaching and location for contrast. In V. Richardson (Ed.), *Handbook of research on teaching* (pp. 333—357). Washington, DC: American Educational Research Association.

Leinhardt, G., & Steele, M. D. (2005). Seeing the complexity of standing to the side: Instructional dialogues. *Cognition and Instruction*, *23*, 87—163.

Lewis, C. (1988). Why and how to learn why: Analysis-based generalization of procedures. *Cognitive Science*, *12*(2), 211—256.

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin and Review*, *6*, 586—597.

Lovett, M. C. (1992). Learning by problem solving versus by examples: The benefits of generating and receiving information. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 956—961). Hillsdale, NJ: Erlbaum.

Mathan, S. A. (2003). Recasting the feedback debate: Benefits of tutoring error detection and correction skills (Unpublished dissertation). Pittsburgh, PA: Carnegie Mellon University.

Mayer, R. E. (1989). Systemic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, *81*, 240–246.

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *The American Psychologist*, *59*, 14—19.

Mayer, R. E., & Anderson, R. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, *83*, 484–490.

Mayer, R. E., & Anderson, R. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, *84*, 444–452.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*(1), 187—198.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, *94*(1), 156—163.

Morgan, P., & Ritter, S. (2002). *An experimental study of the effects of Cognitive Tutor® Algebra I on student knowledge and attitude*. Pittsburgh, PA: Carnegie Learning, Inc. Retrieved from http://www.carnegielearning.com

Mwangi, W., & Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction*, *16*, 173—199.

Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2013). *Effectiveness of Cognitive Tutor Algebra I at scale*. Santa Monica, CA: RAND. Retrieved from http://www.rand.org/pubs/working_papers/WR984.

Pedone, R., Hummel, J. E., & Holyoak, K. J. (2001). The use of diagrams in analogical problem solving. *Memory and Cognition*, *29*, 214—221.

Penner, D. E., Giles, N. D., Lehrer, R., & Schauble, L. (1996). Building functional models: Designing an elbow. *Journal of Research in Science Teaching*, *34*, 125–143.

Perry, M. (1991). Learning and transfer: Instructional conditions and conceptual change. *Cognitive Development*, *6*, 449—468.

Pirolli, P. L., & Anderson, J. R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *39*(2), 240—272.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*, 1—29.

Renkl, A. (2002). Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*, 529—556.

Renkl, A. (2005). The worked-out-example principle in multimedia learning. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 229—246). Cambridge, UK: Cambridge University Press.

Renkl, A. (2011). Instruction based on examples. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 272—295). New York, NY: Routledge.

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, *23*, 90—108.

Ringenberg, M., & VanLehn, K. (2006). Scaffolding problem solving with annotated, worked-out examples to promote deep learning. In K. Ashley & M. Ikeda (Eds.), *Intelligent tutoring systems: Eighth international conference, ITS2006*. (pp. 625—634). Amsterdam, Netherlands: IOS Press.

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*, 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*, 346—362.

Salden, R. J. C. M., Aleven, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science*, *38*, 289—307.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 484—494.

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers and Education*, *46*, 426—445.

Silver, E. A. (1981). Recall of mathematical problem information: Solving related problems. *Journal for Research in Mathematics Education*, *12*, 54—64.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Arithmetic Teacher*, *26*, 9—15.

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19—30). Cambridge, UK: Cambridge University Press.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*, 59—89.

Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251—296.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, *80*, 424—436.

Tobias, S., & Duffy, T. M. (Eds.). (2009). *Constructivist instruction: Success or failure*. New York, NY: Routledge.

Trafton, J. G., & Reiser, B. J. (1993). The contributions of studying examples and solving problems to skill acquisition. In M. Polson (Ed.), *Proceedings of the Fifteenth annual conference of the Cognitive Science Society* (pp. 1017–1022). Hillsdale, NJ: Erlbaum.

Tuovinen, J. E., & Sweller, J. (1999). Comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology*, *91*, 334–341.

Van Dooren, W., De Bock, D., Vleugels, K., & Verschaffel, L. (2010). Just answering … or thinking? Contrasting pupils' solutions and classifications of missing-value world problems. *Mathematical Thinking and Learning: An International Journal*, *12*(1), 20—35.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science*, *32*, 83–98.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, *16*, 154–164.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction*, *18*, 211–222.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*, 1—39.

Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, *43*, 49—64.

Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review*, *22*, 393—409.

Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005). Effects of mathematical word problem solving instruction on students with learning problems. *Journal of Special Education*, *39*, 181—192.