Booth, J.L., Lange, K.E., Koedinger, K.R., & Newton, K.J. (2013). Example problems that improve student learning in algebra: Differentiating between correct and incorrect examples. *Learning and Instruction, 25,* 24-34.

Using Example Problems to Improve Student Learning in Algebra: Differentiating Between Correct and Incorrect Examples

Julie L. Booth[a*]
Karin E. Lange[b]
Kenneth R. Koedinger[c]
Kristie J. Newton[d]

[a]Temple University. 1301 W. Cecil B. Moore Avenue, Philadelphia, PA, 19122, USA.
julie.booth@temple.edu, Phone: 215-204-6223, Fax: 215-204-6013
[b]Temple University. 1301 W. Cecil B. Moore Avenue, Philadelphia, PA, 19122, USA.
karin.lange@temple.edu,
[c]Carnegie Mellon University. 5000 Forbes Avenue, Pittsburgh, PA, 15213, USA.
koedinger@cmu.edu
[d]Temple University. 1301 W. Cecil B. Moore Avenue, Philadelphia, PA, 19122, USA.
kjnewton@temple.edu,

*corresponding author

Funding for this research was provided by the National Science Foundation, Grant Number SBE-0354420 to the Pittsburgh Science of Learning Center (PSLC, http://www.learnlab.org). The research reported here was also supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A100074 to Temple University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Portions of this work were presented at meetings of the Cognitive Science Society, the Eastern Psychological Association, the Cognitive Development Society, and the American Education Research Association. Thanks are due to Golden Valley High School, Central Westmoreland Career and Trade Center, and Shaker Heights Middle School.

ABSTRACT

In a series of two *in vivo* experiments, we examine whether correct and incorrect examples with prompts for self-explanation can be effective for improving students' conceptual understanding and procedural skill in Algebra when combined with guided practice. In Experiment 1, students working with the Algebra I Cognitive Tutor were randomly assigned to complete their unit on solving two-step linear equations with the traditional Tutor program (control) or one of three versions which incorporated examples; results indicate that explaining worked examples during guided practice leads to improved conceptual understanding compared with guided practice alone. In Experiment 2, a more comprehensive battery of conceptual and procedural tests was used to determine which type of examples is most beneficial for improving different facets of student learning. Results suggest that incorrect examples, either alone or in combination with correct examples, may be especially beneficial for fostering conceptual understanding.

Keywords: worked examples; self-explanation; learning from errors; conceptual understanding; Algebra

Using Example Problems to Improve Student Learning in Algebra: Differentiating Between Correct and Incorrect Examples

1. Introduction

Numerous studies have demonstrated that learning is improved when students study worked examples while they practice solving problems in a laboratory setting (e.g., Sweller, 1999; Sweller & Cooper, 1985). Several studies have extended this work to demonstrate that the use of worked examples can be beneficial for learning in real world classrooms over short-term lessons (Carroll, 1994; Ward & Sweller, 1990), over longer-term computerized lessons (Kalyuga, Chandler, Tuovinen, & Sweller, 2001; Kim, Weitz, Heffernan, & Krach, 2009; Schwonke, Wittwer, Aleven, Salden, Krieg, & Renkl, 2007), and, most recently, over long-term traditional lessons (Booth, Koedinger, & Paré-Blagoev, revision under review).

The most robust finding regarding the use of worked examples is that replacing approximately half of the practice problems in a session with fully worked-out examples to study leads to improved procedural knowledge (e.g., Sweller & Cooper, 1985, Ward & Sweller, 1990, etc.). That is, after studying examples, individuals learn to solve problems faster by viewing examples and practicing rather than doing double the practice. This is thought to occur because viewing the examples reduces cognitive load, allowing students to focus on learning and understanding the steps of a problem's solution (Sweller, 1999; Zhu & Simon, 1987). However, variations on the traditional format of worked examples may lead to different types of learning benefits. In the following sections, we describe two such variations: Combination with self-explanation prompts and the study of correct vs. incorrect solutions.

1.1 Worked examples with self-explanation

One common variant is to couple worked examples with prompts for students to explain the information to themselves. Explaining instructional material has been shown to improve learning by forcing students to make their new knowledge explicit (Chi, 2000; Roy & Chi, 2005). Logically, it then follows that asking students to explain examples could further improve their learning over having them simply study examples. Indeed, Renkl, Stark, Gruber, and Mandl (1998) found that including self-explanation prompts with examples of interest calculation problems fosters both near transfer of problem solving skills (i.e., solving the type of problem they practiced) and far transfer (i.e., solving problems that are related, but not isomorphic to those practiced (Haskell, 2001)); Catrambone and Yuasa (2006) also demonstrated that prompting self-explanations yielded greater success at locating the relevant information needed to perform transfer tasks when utilizing computerized databases. Aleven and Koedinger (2002) also demonstrated that adding self-explanation prompts to a computerized tutor leads to increased declarative knowledge in Geometry, and Hilbert, Renkl, Kessler, and Reiss (2008) further suggest that adding self-explanation to worked examples improves students' conceptual knowledge of Geometry. These findings suggest that, at least for mathematical content, the addition of self-explanation allows worked examples to improve students' understanding of the underlying concepts inherent in the problems as well as their ability to carry out the steps they were shown.

1.2 Studying correct vs. incorrect examples

Another dimension on which examples can differ which has received far less research attention is the nature of the solution in the examples. While most worked-example research focuses on the use of correct examples, recent work suggests that asking children to explain a combination of correct and incorrect examples can be even more effective (Durkin & Rittle-

Johnson, 2012; Rittle-Johnson, 2006; Siegler, 2002; Siegler & Chen, 2008); similar results have also been found with older students (Huang, Liu, & Shiu, 2008) and adults (Curry, 2004) . The benefit of explaining errors is twofold. First, it can help students to recognize and accept when they have chosen incorrect procedures, leading to improved procedural knowledge over practice alone or correct examples plus practice (Siegler, 2002). Second, and perhaps more important, it can draw students' attention to the particular features in a problem that make the procedure inappropriate. For instance, consider an example in which the equation $3x - 4 = 5$ is incorrectly solved by subtracting 4 from both sides and resulting in a next problem state of $3x = 1$. By explaining how the procedure led to an incorrect answer, students are forced to both accept that the procedure is wrong, and to notice that the negative sign that precedes the 4 makes it inappropriate to apply the strategy. This can help the student replace faulty conceptual knowledge they have about the meaning of the problem features with correct conceptual knowledge about those features; the acquisition of accurate, deep features with which to represent problem situations is key to building expertise (Chi, Feltovich, & Glaser, 1981). Consistent with this assertion, the combination of correct and incorrect examples has been shown to lead to improvements in both conceptual understanding and procedural skill in Algebra compared with procedural practice alone (Booth et al., revision under review). In this study, the approach was shown to be especially beneficial for minority students (Booth et al., revision under review). However, the inclusion of incorrect examples does not always lead to increased benefit for all learners. For example, Große & Renkl (2007) found that relatively novice learners cannot benefit from incorrect examples when they are expected to locate and identify the error in the example themselves. It makes sense that novice students would have difficulty with this component, given that they likely make many of the mistakes themselves and may not recognize them as incorrect. However, further work is needed to determine if removing this requirement and instead prompting students to explain what happened in the example and why it is incorrect may be beneficial. Further direct testing of correct examples vs. a combination of correct and incorrect examples is thus necessary for varied domains and student populations.

Though research focusing specifically on incorrect examples has only emerged in the past decade, the idea that errors can be effective learning tools is not new. Ohlsson's (1996) theory of learning from errors maintains that individuals choose (perhaps implicitly) between possible actions when solving a problem, but that as initial knowledge is often overly general, beginning learners often choose incorrect options. Ohlsson suggests that in order to improve their task knowledge, learners must first detect an error, identify the overgeneral knowledge that caused the error, and explain what additional conditions or features must be added to the overly general knowledge in order to make it correct.

These ideas are also consistent with overlapping waves theory (Siegler, 1996), which maintains that individuals know and use a variety of strategies which compete with each other for use in any given situation. With improved or increased knowledge, good strategies gradually replace ineffective ones. However, for more efficient change to occur, learners must reject their ineffective strategies, which can only happen if they understand both *that* the procedure is wrong and *why* it is wrong (i.e., which problem features make the strategy inappropriate (Siegler, 2002)). Work in the domain of science on conceptual change is based on a similar premise: inducing cognitive conflict along with providing accurate information can help students to build correct conceptions (e.g., Diakidoy, Kendeou, & Ioannides, 2003; Eryilmaz, 2002).

Collectively, these assertions suggest that the combination of correct and incorrect examples is beneficial because the incorrect examples help to weaken faulty knowledge and force students to attend to critical problem features (which helps them not only to detect and correct errors, but also to consider correct concepts), while the correct examples provide support for constructing correct concepts and procedures, beyond that embedded in traditional instruction. It seems clear that both types of support are necessary, but what if extra support for knowledge construction is achieved through other types of innovative classroom practice? In that case, would it still be optimal to provide a combination of correct and incorrect examples, or would providing incorrect examples alone suffice for improving student learning? In the present study, we test the contribution of correct vs. incorrect examples in the context of such support for knowledge construction--guided problem-solving practice with the Cognitive Tutor, a self-paced intelligent tutor system which provides students with feedback and hints as they practice (Koedinger, Anderson, Hadley, & Mark, 1997).

In a series of two experiments, we examine the relative benefit of explaining correct and incorrect examples alone and in combination for improving conceptual and procedural learning for students beginning Algebra. Algebra 1 is a particularly useful testbed, because it is considered to be a gatekeeper course in which success is thought to be crucial for access to advanced opportunities in mathematics and science. Despite its importance, many students entering Algebra 1 do not have accurate conceptual knowledge about the critical features found in equations, which hinders their ability to succeed in the course. For example, students tend to think that the equals sign is an indicator of where the answer belongs, rather than of balance between the two sides (Baroody & Ginsburg, 1983; Kieran, 1981), that negative signs represent subtraction but do not modify the terms they precede (Vlassis, 2004), and that variables represent a single value (Booth, 1984; Knuth, Stephens, McNeil, & Alibali, 2006; Küchemann, 1978). Such misunderstandings have been found to have a detrimental effect on students' ability to solve equations and to hinder students' ability to learn new Algebraic content (Booth & Koedinger, 2008), and, unfortunately, tend to persist even after targeted classroom instruction (Booth, Koedinger, & Siegler, 2007; Vlassis, 2004). Thus, in order to increase student success in Algebra, it is imperative to utilize more effective methods of helping students to build stronger and more accurate conceptual knowledge, without sacrificing attention to procedural skills. These goals may be accomplished by using a combination of worked examples and self-explanation in the domain of equation-solving, which is fundamental in Algebra 1.

2. The Present Study

The present study was conducted in collaboration with the Pittsburgh Science of Learning Center (PSLC), one of six research centers created by the National Science Foundation (NSF) for the purpose of promoting progress in the learning sciences. The main goals of the PSLC include establishing connections between researchers, educators, and curriculum developers in order to conduct high quality research studies to establish the benefit of instructional practices in real educational settings (Koedinger, Corbett, & Perfetti, 2012; see also http://learnlab.org). Teachers whose classes take part in these *in vivo* experiments are encouraged to attend meetings with researchers to review research project plans and evaluate the potential benefit a particular study may have on student learning or potential difficulties that may arise in classroom implementation. In this setting, researchers, educators, and curriculum developers collaborate to improve study ideas, and in ideal cases, devise new ideas for studies.

There are two research questions for the present study. First, do worked examples with self-explanation improve student learning in Algebra when combined with scaffolded practice solving problems? Previous work in geometry suggests that there is added benefit to explaining examples when embedded within Tutor activities, however, there is no evidence to date on whether students using the Algebra 1 tutor, who are generally younger than geometry students, would benefit similarly. The Cognitive Tutor has already been shown to be more effective for middle school students than traditional instruction (Ritter, Anderson, Koedinger, & Corbett, 2007; What Works Clearinghouse, 2009), thus we are comparing the worked example approach against a more stringent, ecologically valid control group (Corbett & Anderson, 2001). Our specific hypotheses for this research question are that (a) students who explain worked examples will improve more than students who do not (control condition) (Hypothesis 1a), but that these effects will be stronger on the measure of conceptual knowledge (Hypothesis 1b), given that procedural performance is already supported and improved through the use of the Cognitive Tutor (Ritter et al., 2007).

Second, are there differential effects on learning when students explain correct examples, incorrect examples, or a combination thereof? Typical use of examples in classrooms is largely focused on the use of correct examples to show students what they are to be learning how to do. However, as mentioned in section 1 above, there are many reasons to believe that there is value added in explaining incorrect examples (e.g., Siegler & Chen, 2008). Results from such studies support the use of incorrect examples in combination with correct examples; the present study extends this work by evaluating the effectiveness of incorrect examples when they are accompanied by guided practice, but not coupled with correct examples. Our specific hypothesis is that students who explain incorrect examples will improve more than students who explain only correct examples, again particularly for the measure of conceptual understanding (Hypothesis 2).

3. Experiment 1

3.1 Methods

3.1.1 Participants.

Participating in Experiment 1 were 134 high school students in nine Algebra 1 classrooms using the Algebra 1 Cognitive Tutor (Koedinger, Anderson, Hadley, & Mark, 1997). This curriculum is commercially available and used in over 2000 school districts across the United States. Classes which use this curriculum spend a portion of their instructional time in the computer lab using the tutor with students working individually on guided practice problems.

Participating classrooms were recruited from three high schools. We were not permitted to collect demographic data for individual students, but we were able to retrieve ethnicity and SES data at the school level. One school was a west-coast high school in which 16% of students were economically disadvantaged; the ethnic breakdown of attendees was: 43% Caucasian, 8% Black, 37% Hispanic, and 11% Asian. The other two schools were located in the Mid-Atlantic region: a career and trade center for high school students (96% Caucasian, 4% Black), and a suburban high school (94% Caucasian, 5% Black, and 1% Asian). In both of these schools, 30% of attendees were economically disadvantaged. In all cases, students taking regular Algebra 1 (rather than honors or remedial) were tested, thus we expect the demographics of the tested samples to reflect those of the larger school populations.

Students were randomly assigned to one of four conditions: three example-based conditions (1 with correct examples only, 1 with incorrect examples only, and 1 with correct and incorrect examples) and one control condition. Eighteen students were excluded from analysis because they did not complete the posttest, leaving a final sample of 116 students: 86 example-based (30 correct examples only, 31 incorrect examples only, and 25 correct and incorrect examples) and 30 control.

3.1.2 Intervention.

All conditions were situated within the Solving Two-Step Equations unit of the Algebra 1 Cognitive Tutor. Each of the students in the example-based conditions received a total of eight examples interspersed into their practice session, with each example replacing a scheduled guided practice problem. The control condition completed the original tutor unit, which contained only guided practice problems with no examples.

Each of the example-based problems illustrated either a *correct* or an *incorrect* example of solving an equation. Students were then asked to explain both *what* was done in the example and *why* the strategy was either correct or incorrect; both steps are thought to be necessary for improving conceptual understanding (Rittle-Johnson, 2006).

See Figure 1 for sample screenshots of the interface for the examples. In both correct and incorrect examples, students were shown the beginning state of a problem and the state of the problem after the first step; for incorrect examples, the mistakes were always real student errors collected in previous work about the particular features of the problem or about the goal of problem-solving in general (Booth & Koedinger, 2008). Students were told that the step was either correct or incorrect, and were asked to explain first "what" was done in the step, and then "why" it is correct or incorrect. Students built their "what" and "why" explanations from choices of sentence fragments. Students indicated "what" was done by building a sentence from a series of three menus. The choices in the first menu were operations (Added, Subtracted, Multiplied, Divided, and Moved) and once the student selected an operation, the choices in the 2nd and 3rd boxes were dependent on that first choice (until a choice is made in the first box, the other boxes are inactive). Students selected a choice from all three boxes to build an explanation, for instance, "Added 3 to both sides", and then submitted their answer. They had to complete the "what" phase correctly before they could move on to the "why" phase to ensure that they were explaining the appropriate step. For incorrect examples, students indicated “why” the step was wrong by building a sentence from a series of two menus where the choices in the first box indicated whether the overall reason was because it was “illegal” or “legal but not helpful,” and the choices in the second menu were dependent on the choice made in the first menu (e.g., “It was illegal because it combined terms that were not like terms”; “It was legal but not helpful because it did not reduce the number of terms”). For correct examples, students indicated “why” the step was right by using a single menu box to complete the sentences “It is legal because…” and “It is helpful because…”. These sorts of menu-based self-explanations have previously been shown to improve learning and transfer in other mathematical contexts (Aleven & Koedinger, 2002). As with the guided practice problems, feedback and hints were available for all examples.

3.1.3 Measures.

To distinguish meaningful differences in knowledge growth based on the intervention, specific items were designed to measure conceptual knowledge and procedural knowledge. Procedural knowledge, for the purpose of this study, is identified as the *how*, while conceptual knowledge is identified as the *what* and *why*. Procedural knowledge encompasses knowing *how* to complete the task, while conceptual knowledge identifies the important features and their meaning (the *what*), taken together to then understand the appropriateness of the procedures (the *why*) (e.g., Booth, 2011).

<u>3.1.3.1 Conceptual knowledge.</u> To assess students’ conceptual knowledge, we used 54 items that measured understanding of concepts identified in previous research as crucial for algebraic equation solving (e.g., the meaning of the equals sign, the significance of negatives in terms, identification of like terms, etc.; Booth & Koedinger, 2008; Kieran, 1981; Knuth et al.,

2006; Vlassis, 2004; α = .72). The measure was designed to assess a wide variety of concepts related to the content, and combined experimenter-designed items with well-established conceptual items that have been developed by other top researchers in the field (e.g., Alibali, Knuth, Hattikudur, McNeil, & Stephens, 2007; Asquith, Stephens, Knuth, & Alibali, 2007). Sample conceptual items can be found in the left-most column of Table 1.

3.1.3.2 Procedural knowledge. To assess students' procedural knowledge, we used 8 items that required students to carry out procedures to solve problems. Four of the items were isomorphic to those taught during the tutor unit, and four were transfer problems, which included features that students had not yet encountered (e.g., two variable terms in the equation). All items were representative of the types of problems found in Algebra 1 textbooks and taught in Algebra 1 courses; sample items can be found in the middle and right columns of Table 1. Student responses were coded as correct or incorrect. To establish reliability, 30% of the data were coded by a second rater. Inter-rater reliability was sufficient for both isomorphic (κ = .87) and transfer problems (κ = .93).

3.1.4 Procedure.

Participating students were administered a paper-and-pencil test assessing their conceptual and procedural knowledge of algebra. Two parallel forms of the test were created, in which the structure of the problems was held constant across the two tests, but the surface features of the problems (e.g., the numbers and letters) were changed. In order to counterbalance presentation, half of the students in each class were randomly assigned to receive Version A, and the other half took Version B as their pretest. All students completed the pretest at the same time, but worked individually; they took approximately 25 minutes to complete the test. Following completion of the pretest, students logged on to the Cognitive Tutor and automatically received their assigned version of the two-step linear equations unit, which they worked on during all of the scheduled computer lab sessions for their class until the unit was complete. The Tutor provided guided procedural practice solving two-step equations; students received immediate feedback about any errors, and could ask the Tutor for hints if they were unsure of what to do; if assistance was requested of the teacher, it was only given after the student had asked the Tutor for hints, ensuring that the Tutor registered the student's confusion. Students were told to work through the material individually, and at their own pace, until the Tutor had determined that mastery was reached on all requisite skills. This was done in order to maintain the ecological validity of a Cognitive Tutor classroom; had we instead controlled for time on task, students would have gotten varying levels of benefit from the Tutor itself. As soon as possible after each student completed the unit, he or she was administered a posttest[1], which again took approximately 25 minutes. Because students were allowed to complete the unit to mastery, and thus take as long as they needed, the total study time varied for each student; however, most of the students had completed the study within 4 weeks (8 class sessions). There were no significant differences between conditions at pretest ($F(3,115) = 1.75$, *ns*).

3.2 Results

Mean pretest and posttest scores for each condition on each of the three measures can be found in Table 2. To test whether explaining examples provided additional benefit to students when combined with the Cognitive Tutor, we first conducted a MANCOVA for condition

---

[1] Students in 8 of the 9 classrooms were administered the alternate version of the paper-and-pencil test as a posttest (e.g., if they took version A for the pretest, they took version B for the posttest); in one classroom, students were given the same version of the test at pretest and posttest due to teacher error. Scores from this class did not differ significantly from those in the other 8 classes, so data are collapsed across all 9 classes.

(examples vs. no examples) on students' posttest conceptual knowledge and posttest isomorphic and transfer procedural knowledge; composite pretest scores were used as a covariate (see Figure 2). The multivariate effect of condition was significant $F(1,111) = 2.74$, $p < .05$, $\eta_p^2=.07$). Even though students in the experimental condition solved 8 fewer practice problems than those in the control, they performed just as well on isomorphic post-test problems (42% correct vs. 41% in control group, $F(1,113) = 0.03$, $p = .86$, and trended toward better performance on a measure of procedural transfer skill (24% vs. 16% correct, $F(1,113) = 3.43$, $p = .07$, $\eta_p^2=.03$). More importantly, students who received example-based assignments outscored control students on a posttest measure of conceptual knowledge (63% vs. 58% correct, $F(1,113) = 5.28$, $p < .05$, $\eta_p^2=.05$).

Next, to determine whether there were any differences among the three example-based conditions, we conducted a MANCOVA for condition (3: correct only, incorrect only, correct + incorrect) on students' posttest conceptual knowledge, isomorphic procedural knowledge, and transfer procedural knowledge, with composite pretest scores used as a covariate. The multivariate effect did not reach significance, $F(6,162) = 0.96$, $p = .46$, $\eta_p^2=.03$, indicating that there were no differences detected among the three example-based conditions.

3.3 Discussion

Results from this experiment indicate that a combination of examples and guided practice problems is more beneficial for student learning than guided practice alone (Hypothesis 1a). This experience fostered students' conceptual understanding of the features in equations (Hypothesis 1b) without sacrificing their procedural knowledge, which was as good as or perhaps better than that of students in the control group. This study provides evidence that the benefit of worked examples persists, even when combined with a successful, research-based curriculum. It also provides a conceptual replication of the findings of Booth and colleagues (revision under review), demonstrating that the laboratory-proven worked example principle does indeed transfer to classrooms.

Results regarding the relative benefit of correct vs. incorrect examples (Hypothesis 2), however, were not conclusive. In Experiment 1, students in all three example-based conditions performed equally, suggesting that perhaps it doesn't matter what kind of examples a student gets, as long as examples are studied. However, it remains possible that such differences in benefit *do* exist for experience with different types of examples, but they were not detected in Experiment 1.

A more important possibility is that our measures of algebraic competence in Experiment 1 may not have been sufficiently comprehensive. We measured broad conceptual knowledge, isomorphic procedural knowledge, and transfer procedural knowledge, however there are specific facets of each of these types of mathematical knowledge for which the intervention might have had a larger impact. For example, perhaps students in different conditions learn to represent problem features more accurately. Previous work has shown that students do not accurately encode equations, often ignoring or changing the placement of the negative signs, equals signs, and variables in the problem (Booth & Davenport, revision under review). In particular, we expect that viewing incorrect examples might lead students to be more aware of the features in the problem as they try to use them to describe why a procedure is inappropriate, even if it does not lead to differences in their responses on the conceptual knowledge measures of those features. We thus included an additional conceptual knowledge measure of students' encoding of problem features.

Another possibility that we had not previously considered is that the value added of viewing multiple correct examples of the same type (as is done in the correct-only condition) may not lead to greater gains in equation-solving compared with viewing just one example of each type (as is done in the combined condition). Consistent with this assertion, viewing and explaining only one instance has been shown to be sufficient when students' conceptual knowledge is engaged during the instruction (Ahn, Brewer, & Mooney, 1992). However, *comparing* multiple, varied instances of correct examples (Rittle-Johnson & Star, 2007) and *generating* multiple solution methods (Star & Siefert, 2006) have both previously been shown to affect the strategies students choose to use when solving problems on their own, allowing them to become more adept at choosing more efficient strategies when available (e.g., Rittle-Johnson & Star, 2007). We thus included an additional procedural knowledge measure of this problem-solving flexibility, in order to extend this work to providing *self-explanations* of multiple correct examples. . To capitalize on this potential benefit, we also expanded the correct-only condition to include examples of both standard and more unconventional but effective strategies (Star & Seifert, 2006).

Two more minor possibilities are that students didn't receive enough exposure to the different types of examples for there to be a differential effect; in Experiment 1, students received 8 examples over three problem types, meaning that for one of the problem types, students received only two examples, regardless of condition. Exposure to more examples, with better balance across problem types, may be more optimal (recall that in all of the original worked-example studies, Sweller and colleagues utilize a one-to-one ratio of problems to examples (e.g., Sweller & Cooper, 1985)). Another minor possibility may be that the measure of conceptual feature knowledge in Experiment 1 was too broad. In designing the measure, we included items to evaluate student knowledge of each feature that is found in two-step algebraic equations. However, the intervention was focused on improving student knowledge of three features in particular—negative signs, the equals sign, and like terms. A more appropriate conceptual measure that would be more sensitive to changes due to different types of examples may be one that focuses on those three targeted features.

Thus, Experiment 2 was designed to specifically address one overarching research question: What differences in learning exist when students are exposed to the three types of experience with examples? In this experiment, we include more examples in the intervention, balanced across problem type. We also use a more targeted measure of conceptual feature knowledge, a new measure of skill in encoding problem features, and a new measure of flexibility in problem solving, along with measures of isomorphic and transfer equation-solving problems. We again test the hypothesis that students who explain incorrect examples (Combined and Incorrect only conditions) will improve more on measures of conceptual understanding than those who only explain correct examples (Correct only condition) (Hypothesis 2); we also hypothesize that students who explain traditional and unconventional correct examples (Correct only condition) will improve more in flexibility of problem-solving skills than those who do not explain unconventional correct examples (Combined and Incorrect only conditions) (Hypothesis 3).

4 Experiment 2

4.1 Methods

4.1.1 Participants.

Participants in Experiment 2 were recruited from eighth-grade Algebra I classrooms that used the Cognitive Tutor software in an inner-ring suburban school (i.e., a suburb that is located close to a major city) in the mid-western United States, in which 29% of students are classified as low-income. Sixty-four students participated in this study: 29 females (45%) and 35 males (55%). Fifty two percent of the participants were Black, 37% were Caucasian, and 11% identified with multiple races. Students were randomly assigned to one of three example-based conditions (1 with correct examples only (N = 22), 1 with incorrect examples only (N = 21), and 1 with correct and incorrect examples (N = 21)).

4.1.2 Intervention.

The intervention was largely the same as in Experiment 1, with the following changes: First, each condition incorporated 12 examples rather than 8. This was done in order to make sure students had the opportunity to work with the same number of examples in each of the problem types presented by the Tutor (i.e., 4 while solving two-step equations with addition and subtraction, 4 while solving two-step equations with multiplication, and 4 while solving two-step equations with division.)

Second, because exposure to more of the same type of correct examples may not be necessary for learning when students are providing explanations (Ahn et al., 1992), but exposure to varied solution strategies for solving the same type of problem may promote flexibility in problem solving (Rittle-Johnson & Star, 2007), the Correct only condition in Experiment 2 gave students 6 examples solved in the traditional correct method and 6 solved in unconventional but effective ways. To illustrate an unconventional correct example, consider the problem $x/2 – ½ = 3/2$. One could solve the problem traditionally, by adding ½ to both sides and then multiplying both sides by 2 to isolate the variable ($x/2 = 4/2$; $x = 4$). However, it is also feasible to first multiply the whole problem by 2 and then add 1 to find the value of x ($x – 1 = 3$; $x = 4$). This method may not be explicitly taught in school, but is an effective method to make the numbers in the problem easier to manipulate. Thus, for each of the problem types presented by the Tutor (e.g., solving equations with addition and subtraction), students in the Correct only condition received 2 examples using the traditional method for solution and 2 using more innovative methods to find the correct solutions. All correct examples presented to students in the combined group were conventionally solved.

Two minor changes were also made to the interface for the intervention. After analysis of both the problem tasks as well as grade-level vocabulary, two additional options were added to the first pull-down menu to allow students to select "Removed" and "Took the reciprocal of" to describe the mathematics occurring in the steps to solve the equation. The vocabulary of describing why the step was right or wrong was also changed, exchanging the words "legal" and "illegal" for "valid" and "invalid," respectively, to more accurately represent language used by students in the classroom.

4.1.3 Measures.

We utilized two measures of conceptual understanding (feature knowledge and encoding) and two measures of procedural fluency (equation-solving and flexibility). We describe each of the measures in more detail below.

4.1.3.1 Feature knowledge. Feature knowledge questions measured students' understanding of concepts that have been identified in previous research as crucial for success in Algebra; it was comprised of 37 items from the conceptual test in Experiment 1 that measured student knowledge of three critical features: the meaning of the equals sign, negative signs, and like terms ($\alpha = .77$). The percentage of feature knowledge items answered correctly was computed for each student.

4.1.3.2 Encoding. Students' encoding of problem features was measured using a reconstruction task (e.g., McNeil & Alibali, 2004; Matthews & Rittle-Johnson, 2009) in which students were presented with an equation for six seconds and then asked to reconstruct the problem from memory immediately after it disappeared from view. Students completed this task for a series of six equations with different structural formats and different placements of key problem features (e.g., $4x = 9 + 7x – 6$; $p - 5 = -2p + 3$). Student answers were coded for overall correctness, and also in terms of the number and types of errors made on the items (e.g., how many times did they move, drop, or insert a negative sign from an equation; how many times did they switch one numeral with another). For each individual student we computed the percent of equations reconstructed correctly, the total number of errors made on the conceptual features (i.e., those involving the negative sign, equals sign, or variable), and the total number of non-conceptual errors (i.e., those involving letters or numbers). To establish reliability, 30% of the data were coded by a second rater. Inter-rater reliability was sufficient for correct reconstructions ($\kappa = 1.00$), conceptual errors ($\kappa = .80$) and non-conceptual errors ($\kappa = .92$).

4.1.3.3 Equation-solving. As in Experiment 1, we measured students' ability to effectively carry out procedures to solve three isomorphic (i.e., problems that are identical in structure to those they have been trained on; $\alpha = .84$) and three transfer problems (i.e., problems that include additional features or alternate structures from those they have been trained to solve; $\alpha = .84$). For each student, we computed the percent of isomorphic problems answered correctly and the percent of transfer problems answered correctly. Inter-rater reliability (computed on 30% of the data) was sufficient for both isomorphic ($\kappa = .97$) and transfer problems ($\kappa = .92$).

4.1.3.4 Flexibility. We evaluated students' ability to recognize and choose an effective strategy when solving an equation. Students were asked to solve three equations using two different strategies; for these problems, they received a flexibility score that was computed as either a 0, 0.5, or 1 for each problem (α = .70). Receiving a 1 indicated that the student used two different methods to solve the problem, getting the correct answer both times. A score of 0.5 indicated that the student did attempt two different methods, but was not able to come to a correct answer using one of the two methods. A score of 0 indicated that the student was not able to identify two different methods to solve the equation. All items were averaged into a flexibility score, with higher scores indicating that students were able to recognize and use more effective strategies when solving equations.

4.1.3.5 Composite Scores. For each student, we also computed three composite scores: Overall score (percent correct across all feature knowledge, encoding, isomorphic equation-solving, transfer equation-solving, and flexibility items), Conceptual Understanding score (percent correct across feature knowledge and encoding items), and Procedural Fluency score (percent correct across isomorphic equation-solving, transfer equation-solving, and flexibility items). The conceptual understanding and procedural fluency measures were positively correlated ($R(85) = .50$, $p < .001$), and the internal consistency for the composite measure was sufficient (α = .83).

4.1.4 Procedure.

All procedures were identical to those in Experiment 1 except that to avoid confusion about the version of the pretest and posttest that should be given to each student there was only one version of the pencil and paper test; all students received that version as both their pretest and their posttest. There were no significant differences between conditions at pretest ($F(2,83) = .04$, *ns*), and pretest and posttest composite scores were positively correlated ($R(65) = .58$, $p < .001$).

4.2 Results

First, to determine whether there was any effect of condition on students' scores in general, we conducted a 3-level (Condition) ANCOVA on overall scores, controlling for pretest scores (see Figure 3). A trend towards a main effect of condition was found for overall performance ($F(2,64) = 3.063$ $p < 0.10$, $\eta_p^2 = 0.09$). Follow-up pairwise comparisons with Bonferroni correction revealed that students in the Combined condition (M = 63%, SD = 16%) scored higher than students in the Correct only condition (M = 51%, SD = 22%; $p = .05$); no other comparisions were significant.

Mean pretest and posttest scores for each condition on each of the five measures can be found in Table 3. To determine whether there was any effect of condition on the conceptual understanding measures, we conducted a 3-level (Condition) MANCOVA on posttest feature knowledge scores and posttest encoding errors on conceptual features, and posttest encoding errors on non-conceptual features, with overall pretest conceptual understanding scores included as the covariate (see Figure 4, left 2 columns). The multivariate effect of condition was significant, $F(6, 120) = 3.11$, $p < 0.01$, $\eta_p^2 = 0.13$. Significant univariate main effects of condition were found for conceptual features ($F(2,64) = 5.30$, $p < 0.01$, $\eta_p^2 = 0.15$) and conceptual encoding errors ($F(2,64) = 3.56$, $p < 0.05$, $\eta_p^2 = 0.10$). Follow up pairwise comparisons with Bonferroni correction indicated that for the conceptual features measure, students in the Combined condition (M = 72%, SD = 15%) scored significantly higher than students who received the Correct only condition (M = 53%, SD = 26%, $p < 0.01$). In addition, students in the Incorrect only condition made fewer conceptual encoding errors (M = 2.0, SD = 2.1) than students in the Correct only condition (M = 4.4, SD = 4.2; $p < .05$); the comparison between the Combined and Incorrect only conditions was not significant. The univariate main effect of condition on non-conceptual encoding errors was marginally significant $F(2, 64) = 2.55$, $p < 0.10$, $\eta_p^2 = 0.08$), however, follow-up pairwise tests revealed no significant differences between individual conditions.

A parallel 3-level (Condition) MANCOVA was conducted on posttest isomorphic equation-solving, transfer equation-solving, and flexibility scores, with overall pretest procedural fluency scores entered as a covariate (see Figure 4, right 3 columns). The multivariate effect of condition did not reach significance for the procedural fluency measures, $F(6, 120) = 1.38$, $p = .23$, $\eta_p^2 = 0.06$.

5. General Discussion

Experiment 1 indicated that combining guided practice with worked example problems benefited students' conceptual knowledge (Hypotheses 1a and 1b); Experiment 2 measured the impacts of particular types of examples. Results indicated that students performed best after explaining incorrect examples; in particular, students in the Combined condition gained more knowledge than those in the Correct only condition about the conceptual features in the equation, while students who studied only incorrect examples displayed improved encoding of conceptual features in the equations compared with those who only received correct examples (Hypothesis 2).

No differences were found between any of the conditions on any of the procedural measures. For the isomorphic and transfer problems, the combination of guided practice with the Cognitive Tutor and examples yields equal or better performance than the Cognitive Tutor alone (Experiment 1), but the specific types of examples provided may not have much influence on procedural learning when combined with guided practice. In general, the level of performance on the isomorphic and transfer problems was much lower than expected in both studies, especially given that prior research has shown the potential for significant improvement in procedural skill when using the Cognitive Tutor (Ritter et al., 2007). Though all problems were of the types included in the students' textbooks for solving two-step (isomorphic) and multi-step equations (transfer), the easiest possible problems (e.g., ones with all positive numbers and/or numbers that divided or combined neatly) were not represented in the measure. Future work should include simpler problems to more accurately measure the range of learning that occurred in the lesson and determine whether differences in condition emerge for students' improvement in solving such simple problems.

Even more surprising, however, is that students in the Correct only condition—the only condition to include unconventionally solved problems—did not see any significant improvement in procedural fluency when compared to their peers. In particular, we expected that viewing unconventionally solved problems would be crucial for improved flexibility (Hypothesis 3). However, all three groups performed equally on the procedural fluency tasks. One possible reason for this finding is that improved conceptual understanding (as occurred in the Combined and the Incorrect only conditions) may be sufficient to prompt improved flexibility; if students have a better understanding of what the features in the equation actually *mean*, then they may be more likely to construct non-traditional ways of manipulating those features to solve problems. Another possible explanation is that the measure of flexibility in Experiment 2 was focused on students' recognition of multiple effective methods for solving equations rather than their identification of the most effective strategy. A further concern of note is that since only one condition included unconventional examples, it is not clear whether receiving them is not effective or whether receiving them in that particular context is not useful; it is also unknown how much flexibility improves just with the standard Cognitive Tutor curriculum. Future research is necessary to investigate what types and combinations of instruction are sufficient for improving recognition of alternate methods, and whether any of those methods are superior for improving students' strategy choice.

In general, the two conditions that included incorrect examples appear to be more beneficial improving conceptual knowledge than the Correct only condition. This supports and extends previous work in several important ways. First, it provides evidence that the combination of correct and incorrect examples is more beneficial than correct examples alone in the domain of Algebraic equation-solving; previous work in Algebra has focused on correct and incorrect examples of word problems (Curry, 2004). It also confirms the effectiveness of this combination for middle school students (Huang et al., 2008). The present study was also the first to test whether a combination of correct and incorrect examples is more beneficial than *incorrect* examples alone, and suggests that receiving incorrect examples can be beneficial regardless of whether it is paired with correct examples. This finding is especially important to note because when examples are used in classrooms and in textbooks, they are most frequently correctly solved examples. In fact, in our experience, teachers generally seem uncomfortable with the idea of presenting incorrect examples, as they are concerned their students would be confused by them and/or would adopt the demonstrated incorrect strategies for solving problems. Our results strongly suggest that this is not the case, and that students *should* work with incorrect examples as part of their classroom activities.

Our results do not suggest, however, that students can learn solely from explaining incorrect examples. It is important to note that all students saw correct examples, regardless of condition, not only because they are regularly included in textbooks and classroom instruction, but because the correctly completed problems the students produced with the help of the

Cognitive Tutor could also be considered correct examples of sorts. We maintain that students clearly need support for building correct knowledge, however, if that support is coming from another source (e.g., guided practice with feedback), spending additional time on correct examples may not be as important as exposing students to incorrect examples.

Future work should continue to investigate how the structure of incorrect examples influences their benefit for students of varying ability levels. For example, it seems likely that it would be critical to combine incorrect examples with self-explanation rather than promoting aimless examination of errors. Further, the way in which students provide that self-explanation may change their level of benefit from incorrect examples; lower-achieving students in Große and Renkl (2007) were not successful with incorrect examples, but such students in the present study did benefit, perhaps in part because they were supported in constructing reasonable self-explanations through use of the drop-down menu choices. As the benefits of self-explanation are reliant on the learner *generating* the explanation (rather than simply reading a provided explanation; Chi, De Leeuw, Chiu, & Lavancher, 1994) this scaffolding may have allowed students who would not normally be able to construct self-explanations to succeed in the process; affording them benefits for constructing new knowledge (Chi, 2000) and representing the algebra problems (Neuman & Schwarz, 2000).

While results from the present study appear conclusive in supporting the use of incorrect examples in classroom activities, several other important questions are raised from the intricacies of the results. For instance, why did no differences emerge between students in the Combined and the Incorrect-only conditions on any of the tasks? It seems odd that an increased number of incorrect examples wouldn't influence performance, especially since each targeted a different possible misunderstanding about the equation features or solution process. Perhaps as long as students think through ‘enough’ incorrect solutions, there is no added benefit to seeing more. However, then it is strange that the Combined condition—which received ‘enough’ incorrect examples *plus* the same number of correct examples—did not outperform the Incorrect-only condition, even on procedural tasks for which correct examples are frequently shown to improve learning. Perhaps the potential benefit of this combination was minimized because students did not directly compare the two types of examples (Rittle-Johnson & Star, 2007). Further work is needed to investigate these questions, and to determine whether incorrect examples might be more beneficial alone or in combination when they are not also combined with the guided practice.

Another limitation of the present study was that we do not know whether students felt they were learning better from the intervention compared to their typical experience with the tutor, or whether the teacher believed it to be the case. Observations that were conducted in the classrooms suggest that students were able to work well with the intervention and that the lesson went smoothly, however it would be useful to know if the advantages of the example-based tutors are manifest solely in cognitive benefits, or if students’ self-efficacy—or the belief that they can succeed (Bandura, 1986)—changed as a result of working with any of the conditions. Future studies should also examine process data to determine how explaining the examples impacted students’ work with the guided practice problems. For example, did it take them fewer problems to reach mastery? Did they make fewer mistakes when solving problems after explaining correct vs. incorrect examples? It would also be useful to know whether improvements in conceptual understanding as demonstrated on our measures actually yielded noticeable differences in students’ performance in their traditional lessons. All of these factors should be examined in future work.

Nonetheless, though inadequate conceptual knowledge of features in algebraic equations hinders students' success in Algebra (Booth & Koedinger, 2008), the present study demonstrates a successful plan for intervention: Provide examples along with guided practice problems to improve conceptual understanding. Specifically, exposure to incorrect examples which target typical student misconceptions about problem features and problem-solving approaches may be crucial for developing students' conceptual knowledge. Viewing and explaining such incorrect examples may help students to both confront their own misconceptions as well as refine their understanding of the features in the problem; this may be especially so when students are guided to notice critical features in the problems, either by forcing them to explain what changed between steps or by prompting them to answer specific questions about the mistakes in the problem.

Future research should also examine whether the optimal combination of examples differs for students with varied levels of background knowledge; such results would not be unprecedented (Große & Renkl, 2007; Kalyuga, Chandler, & Sweller, 2001). Systematic exploration of such individual differences in future studies is necessary to identify exactly when each type of examples are most beneficial for individual students. This knowledge would have the potential to improve instruction on equation-solving for students in the Cognitive Tutor curriculum, as well as to inform all teachers of example-based strategies that can be used to differentiate algebra instruction for their students.

References

Ahn, W.K, Brewer, W.F., & Mooney, R.J. (1992). Schema acquisition from a single example. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*(2), 391-412. doi:10.1037/0278-7393.18.2.391

Aleven, V., & Koedinger, K.R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26,* 147-179. doi:10.1016/S0364-0213(02)00061-7

Alibali, M. W., Knuth, E. J., Hattikudur, S., McNeil, N. M., & Stephens, A. C. (2007). A longitudinal look at middle-school students' understanding of the equal sign and equivalent equations. *Mathematical Thinking and Learning, 9*, 221-247. doi:10.1080/10986060701360902

Asquith, P. S., Stephens, A. C., Knuth, E. J., & Alibali, M. W. (2007). Middle-school mathematics teachers' knowledge of students' understanding of core algebraic concepts: Equal sign and variable. *Mathematical Thinking and Learning, 9,* 247-270. doi:10.1080/10986060701360910

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Baroody, A. & Ginsburg, H. (1983). The effects of instructions on children's understanding of the equals sign. *The Elementary School Journal, 84,* 199-212. doi: 10.1086/461356

Booth, J.L. (2011). Why can't students get the concept of math? *Perspectives on Language and Literacy, Spring 2011,* 31-35.

Booth, J.L., & Davenport, J.L. (revision under review). The role of problem representation and feature knowledge in algebraic equation-solving. *Journal of Mathematical Behavior*.

Booth, J.L., & Koedinger, K.R. (2008). Key misconceptions in algebraic problem solving. In B.C. Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Cognitive Science Society (pp.571-576). Austin, TX: Cognitive Science Society.

Booth, J.L., Koedinger, K.R., & Paré-Blagoev, E.J. (revision under review). Does practice alone make perfect? Taking example-based assignments for a test drive in real-world classrooms. *Journal of Educational Psychology*.

Booth, J.L., Koedinger, K.R., & Siegler, R.S. (2007). The effect of prior conceptual knowledge on procedural performance and learning in algebra. Poster presented at the 29th annual meeting of the Cognitive Science Society in Nashville, TN.

Booth, L.R. (1984). *Algebra: Children's Strategies and Errors*. Windsor, UK: NFER-Nelson.

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom. *Journal of Educational Psychology, 86(3)*, 360–367. doi:10.1037/0022-0663.86.3.360

Catrambone, R., & Yuasa, M. (2006). Acquisition of procedures: The effects of example elaborations and active learning exercises. *Learning and Instruction, 16, 139*-153. doi:10.1016/j.learninstruc.2006.02.002

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

Chi, M.T.H., De Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477. doi:10.1207/s15516709cog1803_3

Corbett, A.T., & Anderson, J.R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement, and attitudes. In *Proceedings of ACM CHI'2001 Conference on Human Factors in Computing Systems*, 245-252.

Curry, L. (2004). The effects of self-explanations of correct and incorrect solutions on algebra problem-solving performance. In K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society* (pp. 1548). Mahwah, NJ: Erlbaum.

Diakidoy, I.A.N., Kendeou, P., & Ioannides, C. (2003). Reading about energy: The effects of text structure in science learning and conceptual change. *Contemporary Educational Psychology, 28*(3), 335-356. doi:10.1016/S0361-476X(02)00039-5

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*, 206-214. doi:10.1016/j.learninstruc.2011.11.001

Eryılmaz, A. (2002). Effects of conceptual assignments and conceptual change discussions on students' misconceptions and achievement regarding force and motion. *Journal of Research In Science Teaching, 39*(10), 1001–1015. doi:10.1002/tea.10054

Große, C.S. & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning & Instruction, 17*, 617-634. doi:10.1016/j.learninstruc.2007.09.008

Haskell, R. E. (2001). *Transfer of Learning. Cognition, Instruction, and Reasoning.* San Diego: Academic Press.

Hilbert, T. S., Renkl, A., Kessler, S. & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction, 18,* 54-65. doi:10.1016/j.learninstruc.2006.10.008

Huang, T.-H., Liu, Y.-C., & Shiu, C.-Y. (2008). Construction of an online learning system for decimal numbers through the use of cognitive conflict strategy. *Computers & Education, 50,* 61-76. doi:10.1016/j.compedu.2006.03.007

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology*, *21*(1), 5-23. doi:10.1080/01443410124681

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*(3), 579–588. doi:10.1037/0022-0663.93.3.579

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, *12,* 317-326.

Kim, R, Weitz, R., Heffernan, N. & Krach, N. (2009). Tutored Problem Solving vs. "Pure": Worked Examples In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 3121-3126). Austin, TX: Cognitive Science Society.

Knuth, E. J., Stephens, A. C., McNeil, N. M. & Alibali, M.W. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education*, *37*, 297-312.

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., Corbett, A. C., & Perfetti, C. (2012). The Knowledge-Learning-Instruction (KLI) framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36,* 757-798. doi:10.1111/j.1551-6709.2012.01245.x

Küchemann, D. (1978). Children's understanding of numerical variables. *Mathematics in School, 7*, 23-26.

Matthews, P. G. & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations, concepts, and procedures as pedagogical tools. *Journal of Experimental Child Psychology*, *104*, 1-21. doi:10.1016/j.jecp.2008.08.004

McNeil, N. M. & Alibali, M. W. (2004). You'll see what you mean: Students encode equations based on their knowledge of arithmetic. *Cognitive Science*, *28*, 451-466. doi:10.1016/j.cogsci.2003.11.002

Neuman, Y., & Schwarz, B. (2000). Substituting one mystery for another: The role of self-explanations in solving algebra word-problems. *Learning and Instruction*, *10*, 203-220. doi: 10.1016/S0959-4752(99)00027-4

Ohlsson, S. (1996). Learning from error and the design of task environments. *International Journal of Educational Research, 25*(5), 419-448. doi:10.1016/S0883-0355(97)81236-0

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. doi:10.1006/ceps.1997.0959

Ritter, S., Anderson, J.R., Koedinger, K.R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review, 14,* 249-255. doi:10.3758/BF03194060

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, 77, 1–29. doi:10.1111/j.1467-8624.2006.00852.x

Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? An experimental study on learning to solve equations. *Journal of Educational Psychology, 99*(3), 561-574. doi:10.1037/0022-0663.99.3.561

Roy, M. & Chi, M.T.H. (2005). Self-explanation in a multi-media context. In R. Mayer (Ed.), *Cambridge Handbook of Multimedia Learning* (pp. 271-286). Cambridge Press.

Schwonke, R., Wittwer, J., Aleven, V., Salden, R.J.C.M., Krieg, C., and Renkl, A. (2007). Can tutored problem solving benefit from faded worked-out examples? Paper presented at The European Cognitive Science Conference, Delphi, Greece.

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

Siegler, R.S. (2002). Microgenetic studies of self-explanations. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31-58). New York: Cambridge University.

Siegler, R. S., & Chen, Z. (2008). Differentiation and integration: Guiding principles for analyzing cognitive change. *Developmental Science, 11*, 433-448. doi:10.1111/j.1467-7687.2008.00689.x

Star, J. R., & Seifert, C. (2006). The development of flexibility in equation solving. *Contemporary Educational Psychology, 31*, 280-300. doi:10.1016/j.cedpsych.2005.08.001

Sweller, J. (1999). *Instructional design in technical areas.* Victoria, Australia: Australian Council for Education Press.

Sweller, J., & Cooper, G. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59-89. doi:10.1207/s1532690xci0201_3

Vlassis, J. (2004). Making sense of the minus sign or becoming flexible in ‘negativity’. *Learning and Instruction, 14,* 469-484. doi:10.1016/j.learninstruc.2004.06.012

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*(1), 1-39. doi:10.1207/s1532690xci0701_1

What Works Clearinghouse (2009). *Middle School Math: Cognitive Tutor Algebra 1.*

Zhu, X., & Simon, H.A. (1987). Learning mathematics from examples and by doing. *Cognition and Instruction*, 4, 137-166. doi:10.1207/s1532690xci0403_1

Table 1: Sample assessment items for Experiment 1.

Conceptual	Procedural Isomorphic	Procedural Transfer
State whether each of the following is equal to -4x + 3: a. 4x + 3 Yes No b. 3 – 4x Yes No c. 4x – 3 Yes No d. 3 + (-4x) Yes No e. 3 + 4x Yes No	-4x + 5 = 8	6 + 3y = -5y + 8
If 10x – 12 = 17 is true, state whether each of the following must also be true: a. 10x – 12 + 12 = 17 + 12 Yes No b. x – 2 = 17 Yes No c. 10x = 29 Yes No d. 10x = 17 Yes No e. 10x – 10 – 12 – 10 = 17 Yes No f. 10x – 12 + 12 = 17 Yes No	$\frac{6}{b} = 9$	$5 = \frac{8}{t} - 1$
State whether each of the following is a like term for 8c: a. -5 Yes No b. 8 Yes No c. 6c Yes No d. 3d Yes No e. 5(c+1) Yes No f. -4c Yes No	$3 = -\frac{d}{7}$	$7 - \frac{v}{8} = 2$

Table 2. Mean pretest and posttest scores on each measure in Experiment 1 for students in each condition

	Conceptual Knowledge		**Isomorphic Procedural**		**Transfer Procedural**	
	Pre	Post	Pre	Post	Pre	Post
Control	.62 (.09)	.60 (.11)	.43 (.27)	.44 (.27)	.16 (.19)	.18 (.22)
Correct Only	.53 (.15)	.62 (.14)	.33 (.27)	.45 (.32)	.23 (.26)	.28 (.30)
Incorrect Only	.59 (.13)	.65 (.10)	.32 (.26)	.40 (.29)	.18 (.26)	.24 (.23)
Combined	.59 (.11)	.60 (.14)	.35 (.25)	.39 (.34)	.20 (.23)	.24 (.30)

Note: Mean (SD).

Table 3. Mean pretest and posttest scores on each measure in Experiment 2 for students in each condition

	Encoding Conc Errs		**Conceptual Features**		**Isomorphic Procedural**		**Transfer Procedural**		**Flexibility**	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Correct Only	5.04 (3.18)	4.48 (4.20)	.67 (.16)	.52* (.26)	.30 (.28)	.28 (.26)	.13 (.22)	.19 (.30)	.15 (.26)	.13 (.29)
Incorrect Only	3.67 (2.59)	1.90* (2.10)	.67 (.19)	.63 (.26)	.51 (.39)	.40 (.33)	.32 (.32)	.30 (.38)	.22 (.30)	.15 (.31)
Combined	4.38 (2.52)	2.90* (2.55)	.66 (.16)	.71* (.17)	.30 (.35)	.29 (.30)	.22 (.34)	.32 (.32)	.14 (.33)	.18 (.32)

Note: Mean (SD). *Posttest score significantly different from pretest score at $p<.05$

Figure Captions

Figure 1: Screenshots of the interface for correct and incorrect examples in Experiment 1.

Figure 2: Conceptual, procedural isomorphic, and procedural transfer scores for example-based and control students in Experiment 1.

Figure 3: Overall posttest scores (adjusted for pretest scores) for students in each condition in Experiment 2.

Figure 4: Encoding, Feature Knowledge, Isomorphic Equation-Solving, Transfer Equation-Solving, and Flexibility posttest scores (adjusted for pretest performance) by condition in Experiment 2.