

Effect of worked examples and Cognitive Tutor training on constructing equations

Stephen K. Reed · Albert Corbett · Bob Hoffman · Angela Wagner · Ben MacLaren

Received: 7 June 2011 / Accepted: 6 January 2012 / Published online: 20 January 2012
© Springer Science+Business Media B.V. 2012

Abstract Algebra students studied either static-table, static-graphics, or interactive-graphics instructional worked examples that alternated with Algebra Cognitive Tutor practice problems. A control group did not study worked examples but solved both the instructional and practice problems on the Cognitive Tutor (CT). Students in the control group requested fewer hints and made fewer errors on the CT practice problems but required more learning time on the instructional examples. There was no difference among the four groups in constructing equations on a paper-and-pencil posttest or on a delayed test that included training and transfer problems. However, students who studied worked examples with a table were best at identifying the meaning of the equation components. The concept of transfer-appropriate processing (the overlap between instructional task and assessment task) aided our interpretation of the findings. Although the CT had a short-term effect on reducing errors and hint requests on CT practice problems, the worked examples were as effective on delayed paper-and-pencil tests. The subsequent construction of a new module for the Animation Tutor (Reed and Hoffman, Animation Tutor: Mixtures. Instructional software, 2011) used both the interactive-graphics and static-table worked examples to take advantage of the complementary strengths of different representations (Ainsworth, Learn Instr 16:183–198, 2006).

Keywords Cognitive Tutor · Transfer-appropriate processing · Worked examples · Word problems

A broad ongoing debate in instructional design concerns the relative effectiveness of minimally guided instruction versus direct instructional guidance. Kirschner et al. (2006) argue that direct instruction improves learning outcomes by reducing the students' need to search for solutions. They argue that such search activities tie up limited working memory

S. K. Reed (✉) · B. Hoffman
CRMSE, San Diego State University, 6475 Alvarado Road, San Diego, CA 92182, USA
e-mail: sreed@sunstroke.sdsu.edu

A. Corbett · A. Wagner · B. MacLaren
Carnegie Mellon University, Pittsburgh, PA, USA

capacity without directly leading to learning. Hmelo-Silver et al. (2007) respond to Kirschner et al.'s (2006) claim by arguing that inquiry learning is often effective because it is supported by scaffolding that does offer guidance.

The use of unassisted problem solving in math and science curricula is a microcosm of this broader debate and one in which the interplay of minimally guided problem solving and direct instruction in the form of worked examples has been extensively studied. In a seminal study Sweller and Cooper (1985) showed that, for solving algebra equations, interleaving worked examples with student problem solving generally led to greater acquisition (as measured by posttest problem solving) in less learning time. This result has been extensively replicated for novice students across a wide variety of mathematics and science domains (Paas 1992; Paas and van Merriënboer 1994; Pashler et al. 2007; Tarmizi and Sweller 1988; van Gog et al. 2004). Interleaved worked examples can also lead to better transfer to solving novel problems than unassisted problem solving (Cooper and Sweller 1987; Ward and Sweller 1990). An exception is the “expertise reversal” effect that occurs when experienced students learn more from unassisted problem solving than from interleaved worked examples (Kalyuga et al. 2003, 2001a, b).

Worked-example solutions and unassisted problem solving lie at the guided and unguided extremes of what Koedinger and Alevan (2007) call the “assistance dilemma”—how much information to provide students to optimize student learning. They note that Cognitive Tutors (CTs) provide an intermediate level of assistance. In this environment students generate problem solutions and the tutors typically provide step-by-step accuracy feedback, requiring the student to remain on a successful solution path. Upon request a CT will also provide advice on the student's next step so the student always reaches a successful conclusion (Anderson et al. 1995). The result is that CTs have yielded dramatic improvements in student learning compared to unassisted problem solving (Anderson et al. 1995; Ritter et al. 2007).

The effectiveness of the CTs raises the question of whether the impact of interleaved worked examples generalizes to problem solving that is supported by an effective tutor. Problem solving with step-by-step feedback and advice as provided by CTs is also expected to decrease extraneous cognitive load (Paas et al. 2004) and yield better learning than unassisted problem solving with less time on task (Corbett and Anderson 2001). Ten experiments have compared intelligent-tutor assisted problem solving to either interleaved worked examples or “faded” worked examples (Renkl and Atkinson 2003) in CTs for stoichiometry (McLaren et al. 2008), algebra equation solving (Anthony 2008), geometry problem solving (Salden et al. 2010; Schwonke et al. 2009, 2011) and in a closely related ASSISTment intelligent tutor for topics in introductory statistics (Weitz et al. 2010).

These comparisons yield two consistent results. The six studies that report learning times consistently replicate a major finding in the classic comparisons of worked examples and unassisted problem solving, that learning is more efficient when worked examples are incorporated with problem solving. Even with the feedback and advice available in assisted problem solving, students completed the same set of problems more quickly with worked examples. However, these studies generally failed to replicate the finding from unassisted problem solving that incorporating worked examples resulted in improved problem solving on immediate tests. Seven of the studies found no difference in immediate problem solving, one study (Schwonke et al. 2011) found that assisted problem solving alone was superior to incorporating worked examples for simple principles, but found no differences for difficult principles, while two experiments (Salden et al. 2010) found some benefits for one type of faded worked examples on immediate problem solving.

The algebra and geometry studies found mixed results for the third fundamental issue raised in the classic worked examples literature—whether worked examples yield deeper understanding of the problem-solving domain. Anthony (2008) found greater retention of problem solving knowledge in her interleaved worked example condition and Salden et al. (2010) found greater retention of problem-solving knowledge for one type of faded worked examples. The first experiment in Schwonke et al. (2009) obtained greater transfer to conceptually related problems in the worked-example condition but that result was not replicated in their second experiment nor in Schwonke et al. (2011).

We continue the investigation of this topic by conducting an exploratory study of linear algebra modeling within the Algebra CT by following the recommended procedure of interweaving worked examples with problem solving (Pashler et al. 2007). After studying each worked example of a word problem students used the CT to solve an equivalent practice problem that had the same story content (either interest or ore) and the same equation structure as the previous worked example. A control group solved the instructional problems on the CT rather than study worked examples. Unlike when solving problems using paper and pencil, students in this group received hints and feedback to evaluate whether assisted problem solving would be more effective than worked examples.

Purpose

Our study continued the investigation of the relative merits of worked examples versus CT assisted problem solving but included several changes from previous studies. First, our study investigates problem-solving tasks in an Algebra CT in which students generated algebraic expressions to model real-world problem situations. Earlier studies, in contrast, focused on the analysis and manipulation of explicit formal representations (e.g., geometric problem solving or algebra equation solving). In our study, students learned to write equations to model “mixture” problems such as

You have an American Express credit card with a balance of \$715 at an 11% interest rate and a Visa credit card with a 15% interest rate. If you pay a total of \$165 in annual interest, what is the balance on your Visa card?

The learning goal was to construct a symbolic model of the situation that could be used to solve the problem, such as

$$(0.11 \times \$715) + (0.15 \times V) = \$165$$

Extending the investigation of worked examples to the Algebra CT is important because of its widespread use in algebra classrooms.

A second departure is that previous studies used worked examples that were implemented in, and fully isomorphic to, the tutor-supported problem-solving task but the algebra modeling task evaluates the relative effectiveness of different types of worked examples. In our study, a table is included in worked examples for the static table instructional group and in CT problem solving to scaffold the relationship between real-world problem referents and the hierarchical components of the mathematical models. In contrast, graphical representations were developed for the static graphics and interactive graphics instructional groups to illustrate the quantitative relationships among the problem referents. Students in CT Algebra courses were randomly assigned to one of the four instructional groups; three groups distinguished by the format of their interleaved worked examples (static table, static graphics, or interactive graphics) and a fourth CT control

group who solved all problems on the Algebra CT. We did not create an interactive table group because the CT includes interactive tables and we designed the worked examples to differ from CT instruction.

The static table group studied text-based explanations of the solutions that included a table for linking components in the equation to the concepts in the problem. Such solutions are typically provided in textbooks and in research on word problems (Reed 1999). The text-based format is also similar to the instructional guidance provided by the CT. The two graphics groups studied either static or interactive bar-graph representations of the solutions to encourage visual thinking (Reed 2010). Emphasis on perception and action (manipulation in the interactive graphics condition) is consistent with theories of grounded cognition in which perception and action are central, rather than peripheral, aspects of cognition (Barsalou 1999; Wilson 2002).

A third departure from previous studies is our use of standard, rather than enhanced, worked examples that had used fading (Salden et al. 2010; Schwonke et al. 2009) or self-explanations (Berthold et al. 2009). Although these enhancements are typically effective, they raise questions about the effectiveness of combining standard (nonenhanced) examples with the CT. Our use of standard worked examples also simplifies comparisons of the static table, static graphics, and interactive graphics instruction because self-explanations and fading would differ across these different instructional conditions.

Finally, because previous studies have yielded inconsistent results on robust learning, in this study we employed three distinct measures of students' depth of understanding of the problem solving domain: retention, transfer to novel but related problems, and an assessment of students' ability to map symbols in an equation onto their real-world referents.

Instructional conditions

Figure 1 shows a worked example for the static table group. This instruction uses a table to represent and label the components of the equation before the components are combined to form an equation. Figure 2 shows a worked example for the static graphics group in which stacks of money represent credit card debt and darker colors represent amount of interest as a percentage of the debt. The numbers beneath the stacks are used to form the equation.

The instruction for the interactive graphics group was the same as for the static graphics group except it attempts to take advantage of the enactment effect (Engelkamp 1998; Reed 2008) by requiring students to manipulate the bars. Construction for the arithmetic word problems required clicking on the interest in the first bar and dragging a copy to the location of the Total Interest bar; then clicking on interest in the second bar and placing a copy on top of the copy from the first bar. For algebra word problems students raised and lowered the bar that represented a variable to observe how changing the variable influenced numbers in the equation and the height of other bars. The goal was to make the difficult concept of a variable (MacGregor and Stacey 1997; Philipp 1992) directly observable by showing how changing its value would impact other values in the equation.

The CT activities for the instructional and practice problems were developed with the Cognitive Tutor Authoring tool (Aleven et al. 2009). Figure 3 shows that the CT problems employ the same table as the static table worked examples, which is intended to scaffold the relationship between the problem referents and the symbols in the equation. As in all CTs, step-by-step accuracy feedback was provided and advice was available upon request.

Solutions

You have a MasterCard with a balance of \$532 at a 21% interest rate. You also have a Visa credit card with a balance of \$841 at a 16% interest rate. How much money are you paying in total annual interest?

The amount of interest for each card is found by multiplying the interest rate percentage by the balance amount. The amount of interest owed for the MasterCard is $0.21 \times \$532$. The first line in the table shows this amount.

The amount of interest owed for the Visa card is $0.16 \times \$841$. The second line in the table shows this amount.

The total annual interest equals the interest for the Mastercard and the interest for the Visa card.

Credit Card	Balance (\$)	Interest Rate	Amount of Interest Owed(\$)
Mastercard	532	21%	.21 x 532
Visa	841	16%	.16 x 841

The sum of these two amounts therefore equals:

$$(.21 \times \$532) + (.16 \times \$841) = \text{Interest}$$

Continue

11

Fig. 1 Worked example for the static table group




Solutions

You have a MasterCard with a balance of \$532 at a 21% interest rate. You also have a Visa credit card with a balance of \$841 at a 16% interest rate. How much money are you paying in total annual interest?

The amount of interest for each card is found by multiplying the interest rate percentage by the balance amount. The amount of interest owed for the MasterCard is $0.21 \times \$532$. The darker green part of the stack shows this amount.

The amount of interest owed for the Visa card is $0.16 \times \$841$. The darker green part of the stack shows this amount.

The total annual interest equals the interest for the Mastercard and the interest for the Visa card.


+

=


\$532

21% interest

\$841

16% interest

\$?

Total Interest

The sum of these two amounts therefore equals:

$$(.21 \times \$532) + (.16 \times \$841) = \text{Interest}$$

Continue

11

Fig. 2 Worked example for the static graphics group

Connected • AuthorTimeTutoring • v.2.3.61

Hint

You have an American Express credit card with a balance of \$715 at an 11% interest rate and a Visa credit card with a 15% interest rate. If you pay a total of \$165 in annual interest, what is the balance on your Visa card?

Unknown = Variable =

	Balance	Interest Rate	Interest Owed
American Express	715	11%	0.11*715
Visa	B	15%	0.15B
Total	---		165

Equation:

Fig. 3 Cognitive Tutor work space

Theoretical comparisons

The many variables that influence learning can make it difficult to make predictions about the relative effectiveness of different instructional conditions, particularly when the instruction includes multiple representations (Ainsworth 2006). Ainsworth's DeFT framework proposes that such predictions need to be based on the *design* parameters that are unique to learning with multiple representations, the *functions* that multiple representations serve in supporting learning, and the cognitive *tasks* that are undertaken by a learner. As a result, different instructional approaches, including the ones investigated in this study, seldom differ along a single dimension.

We propose that the four different types of instruction—CT problem solving, static table, static graphics, and interactive graphics worked examples—serve different design goals in supporting learning (Table 1). The five goals are (1) provide declarative knowledge, (2) organize information in a table, (3) organize information in a diagram, (4) encourage interactive learning, and (5) create multiple memory codes.

Provide declarative knowledge

The initial design of the CTs was based on the ACT* theory of learning in which complex skills are decomposed into a set of production rules (Anderson 1990). However, the theory assumed that all knowledge begins in a declarative representation that is typically acquired from instructions and/or by studying worked examples. Solving problems converts the declarative knowledge into efficient procedures for solving specific problems. Although initial versions of the theory emphasized learning declarative knowledge from instructions,

Table 1 Design features of instruction

Learning conditions	Declarative knowledge	Table organization	Diagram organization	Interactive learning	Multiple memories
Cognitive Tutor		×		×	
Static table	×	×			
Static graphics	×		×		×
Interactive graphics	×		×	×	×

more recent versions emphasize learning declarative knowledge from worked examples (Anderson and Fincham 1994). Providing worked examples before students confront comparable problems on the CT is therefore consistent with the ACT* theory of learning.

Organize information in a table

Worked examples of word problems often include a table that links numbers and symbols to the concepts described in the text. For instance, the table in Fig. 1 shows that the amount of interest owed on the Mastercard is $.21 \times \$532$. Reed and Ettinger (1987) found that including completed tables in worked examples helped students construct equations. However, requiring students to enter quantities into the tables was not helpful because students often entered incorrect values.

The CT requires that students enter values into tables (Fig. 3) but provides feedback that enables them to complete the table correctly. This difference between the two conditions allows us to compare completed tables in the static-tables worked examples with the constructive feedback provided by the CT.

Organize information in a diagram

The static and interactive graphics conditions use diagrams instead of tables to organize the problem information. Displaying quantities in rectangles has traditionally played an important role in helping students represent quantities and relationships among quantities in word problems (Fuson and Willis 1989), including part/whole relations (Gamo et al. 2010; Ng and Lee 2009). The bar diagram in Fig. 2 shows part–whole relations in each stack of money by representing the amount of interest as a percentage of the amount owed on each credit card. It also shows that the total interest is the sum of the interest owed on both credit cards.

Goldstone et al. (2009) have advocated that scientific and mathematical reasoning should be grounded in perceptual processing. They refer to the research of Goldstone and Son (2005) as demonstrating the perceptual basis of successful transfer across two complex visual simulations based on the same (isomorphic) principles. Our instructional and practice problems included ore problems that are isomorphic to the credit card problems. Isomorphic problems have the identical equation structure but different story content, as shown by the corresponding interest and ore types in Appendix 1. The bar graph representation was the same for the ore problems in which the amount of iron or gold was shown as a percentage of the mined ore. The same diagrammatic depiction of the interest and ore problems may facilitate transfer between the two sets of problems.

Encourage interactive learning

Chi (2009) proposes a conceptual framework that she uses to support the hypothesis that interactive learning is more effective than constructive, active, and passive learning. Interactive learning includes feedback and guidance provided by a computer system or pedagogical agent. Moreno and Mayer (2007) introduced a special issue on interactive learning environments by referring to five types of interactivity. One type—manipulating—is a central feature of the interactive graphics condition.

Constructing the Total Interest bar for the arithmetic problems and adjusting the unknown variable for the algebra problems, however, does not provide the extensive feedback that is a central feature of the CT. The tutor instruction provides the most

interactive learning followed by the interactive-graphics condition. The static-graphics and static-table conditions do not allow for interactivity.

Create multiple memory codes

An advantage of the two graphics conditions is that the equation is shown both visually and symbolically. According to dual coding theory (Paivio 1986) two memory codes should provide a better memory of the solution. Mayer's (2001) research provides extensive support for the multimedia principle that students learn more from pictures and words than from words alone.

An advantage of the interactive graphics condition over the static graphics condition is that manipulation provides a third (motor) code that can also enhance memory (Engelkamp 1998; Reed 2006, 2008). There is extensive evidence provided by Engelkamp and others that enactment aids recall. The interactive graphics condition supports visual, verbal, and motor codes; the static graphics condition supports visual and verbal codes; and the static table and tutor conditions support verbal codes. Additional memory codes should be particularly advantageous during delayed testing.

In conclusion, the four instructional conditions have different strengths summarized in Table 1 that make it difficult to make predictions about their relative effectiveness. This study is therefore exploratory.

Method

Participants and context

A total of 128 students enrolled in CT Algebra courses in three Pittsburgh area high schools participated in the project. Students completed the study activities as part of their sessions in their computer labs that ran the Algebra CT software (Koedinger et al. 1997). The study involved mixture problems that can be represented by general linear models of the form $ax + by = c$ that are discussed in Units 15–16 of the Algebra I CT curriculum.

Students at two of the schools were enrolled in CT Algebra II. This course begins with a two-chapter review of linear functions, which these students had completed shortly before participating in this study in the late fall. Students in these courses were in the CT lab two class periods per week. Students at the third school were enrolled in the second year of a two-year Algebra I sequence. Students in this course had also completed work on linear functions and were in the CT lab one double-period per week.

Design and procedure

The study consisted of three consecutive computer lab sessions that were typically separated by either several days or 1 week. Students received eight arithmetic (mixture) problems during session 1 that were used to help prepare them (Moreno and Mayer 2007) for the eight algebra word problems in session 2. Session 3 contained a variety of test problems to measure robust learning.

We based our decision to begin with arithmetic, rather than algebra, problems in session 1 on the assumption that arithmetic problems are typically easier than algebra problems, as proposed in Heffernan and Koedinger's (1998) developmental model of algebra

symbolization. Performance of high school students on a variety of arithmetic and algebra problems supported their model. The arithmetic problems therefore helped prepare students for the more demanding algebra problems in session 2.

The students in each of the three schools were randomly assigned to one of the four instructional conditions. In each of the first two sessions, students in the three worked example conditions studied worked example solutions of the four odd-numbered instructional problems and solved the four even-numbered practice problems with the CTs. Students in the CT condition solved all eight problems using the CT. The even-numbered CT practice problems, which were common to all four groups, served both as learning experiences for all the students and as evaluations of the effectiveness of the four different learning activities by determining how much assistance was required for further learning. Both instructional and practice problems were presented on a computer to ensure fidelity of treatment across training conditions.

Learning materials

Four types of mixture problems were developed, two arithmetic types, for use in session 1, and two algebraic types for use in session 2. Two pairs of equivalent instructional–practice problems were developed for each of the four problem types, for a total of 16 problems. One pair of each type involved interest payments on two credit cards, as displayed in the figures. The other pair of each type consisted of mining problems, which involved extracting metal from two ores of different quality (metal content). In all 16 problems, the goal is to generate an equation as a mathematical model of the problem situation (but not to solve the resulting equation). Appendix 1 shows the initial (instructional) problems for each of the eight problem pairs. The four problems of each type were presented in succession, in the order shown in Appendix 1. Within each type, students completed the pair of equivalent interest problems first, followed by the pair of equivalent ore problems.

We selected both interest and ore problems to provide students with the opportunity to apply general linear models in two different domains. Reed (1987) found that students had difficulty in both mapping quantities and constructing equations across isomorphic versions of mixture problems, including ore and interest problems. We therefore hoped that instruction on interest problems would help students construct equations for ore problems.

The displayed graphic solutions follow two principles of multimedia design based on the research of Sweller (2003) and Mayer (2001; Moreno and Mayer 2007). The first is the proximity principle that different media be closely integrated in space. Text explanations are therefore placed above each of the bars and the equation appears directly below the bars (Fig. 2). The second principle, minimize cognitive load, is achieved by presenting the solution in three segments. After seeing the problem statement, students successively press the Continue arrow to see the (1) first bar with explanation, (2) second bar with explanation, and (3) third bar with explanation and equation. The text above the table in the static-table worked examples also occurred in three segments (Fig. 1). Gerjets et al. (2004) presented evidence showing that such modular presentation of worked examples improves learning.

During session 1, students in the interactive graphics group clicked on and moved copies of the interest bars (the darker green parts of the stacks) to construct the Total Interest bar on the right side of the equation. The first interest bar and its copy have a red border when clicked and the second interest bar and its copy have a blue border when clicked. The borders enable students to distinguish the two quantities in the Total Interest bar and relate them to their referents on the left side of the equation. Clicking on bars also changes the color of terms in the equation (red for the first term and blue for the second

term) to dynamically link the graphics to the equation (Berthold and Renkl 2009). Students could not advance in the instructional sequence until they completed these manipulations.

During session 2, students in the interactive graphics group changed the value of the unknown variable by raising and lowering the height of this bar for each of the four worked examples. Varying this bar enabled them to see changes in the height of other bars and in the values in the equation. Only the height of the bar on the right side of the equation (total interest or ore) changed for the first two problems but both of the other bars changed for the last two problems because the unknown variable occurs twice in these equations (such as D and $\$1405 - D$). Students were instructed to restore the bar to its initial height following this exploration.

We implemented the graphics worked examples by integrating the graphics capabilities of the Animation Tutor (Reed 2005) with the CT. This software integration took longer than anticipated so we were only able to provide interactive graphics for the first worked example in session 1. The interactive graphics group studied static graphics for the remaining three worked examples so we are unable to draw conclusions about interactive graphics for the arithmetic word problems. In session 2, students in the interactive graphics group adjusted the bar representing the unknown variable in all four worked examples so we are able to evaluate the effect of manipulations for algebra word problems. Readers can perform these manipulations by downloading the Mixtures module from the Animation Tutor web site (<http://www.sci.sdsu.edu/mathtutor/index.htm>).

Test materials

Four measures of student learning from the 16 on-line problems were developed, an immediate test of students' algebra problem solving knowledge and three tests of the "robustness" of student learning. These three robust learning tests included (a) a retention test of problem solving knowledge, (b) a transfer test of problem solving knowledge, and (c) a test of students' explanations of mathematical model structure.

Session-2 problem-solving test

This paper-and-pencil test consisted of one interest and one ore problem that were equivalent to the two types of algebra problems presented in the Session-2 on-line environments. Students were asked to generate an equation to model the situation in each problem. Two test forms were developed so each form could serve as the pretest for half the students who then switched to the other form in the posttest.

Session-3 retention test

This paper-and-pencil test consisted of four problems that were equivalent to the four types of problems students had encountered in the on-line learning environments. Students were asked to generate an equation to model the situation for two interest and two ore problems.

Session-3 transfer test

This paper-and-pencil test consisted of one arithmetic interest problem and one algebra ore problem with novel situations. Students were asked to model these situations with equations that also had novel structures.

Session-3 model component descriptions

The CT Model Analysis Tool (Corbett et al. 2000, 2003, 2007) was employed to ask students to explain the structure of the arithmetic and algebraic models. As displayed in Appendix 2, each problem presents a problem description and a mathematical model of the situation. Students select entries from menus to describe what each hierarchical component of the symbolic model represents in the problem situation. As in all CTs, students received feedback on each problem step, could request advice on each step, and were required to complete a correct solution to the problem. Four Model Analysis problems corresponded to the four problem types that students had solved during the prior two sessions.

Results

All tasks, in both learning and testing, required constructing an equation so correct solutions refer to correct equations rather than to correct numerical answers. We excluded 4 students from the data analysis for missing the second session and seven students because they talked to others as they worked on the problems. The excluded students included three from the CT group, one from the ST group, three from the SG group, and four from the IG group. Of the remaining 117 students, there were 29 in the CT condition, 29 in the static table condition, 28 in the static graphics condition and 31 in the interactive graphics condition. Some of the evaluations are based on fewer than the remaining 117 students because of missing students or missing data. We report the sample size for each analysis. The total number of missing cases, including the 11 excluded from all analyses, was 15 from the CT group, 11 from the ST group, 14 from the SG group, and 15 from the IG group. Missing cases are therefore fairly uniformly distributed across groups.

Task times (sessions 1 and 2)

Each of the learning sequences in sessions 1 and 2 consisted of four pairs of problems in which an initial worked example (or CT problem in the control condition) was followed by an equivalent CT practice problem. The instructional software did not record when students began to study the first worked example so we do not know the amount of time they spent working on the first pair of problems. We were able to record the amount of time spent on each of the six problems in the second, third, and fourth pairs. We excluded 13 students (four CT, three static table, three static graphics, three interactive graphics) from analyses of the first session and 16 students (six CT, two static table, four static graphics, four interactive graphics) from analyses of the second session because of missing data.

An ANOVA of the learning times for the three pairs of arithmetic problems during the first session revealed significant differences across the four instructional groups, $F(3, 100) = 6.88$, $p < .001$, partial $\eta^2 = .171$. Bonferroni comparisons revealed that the CT group differed from each of the three worked example groups at a $p = .02$ level or lower, and the three worked example groups did not differ from each other. The CT group took more time per problem than the static table group ($M = 33.27$, $SD = 10.86$), static graphics group ($M = 36.88$, $SD = 11.07$), and the interactive graphics group ($M = 46.83$, $SD = 10.86$). Instruction and practice problems followed immediately after the conclusion of the previous problem so the total time spent on these problems is the total instructional time.

These times, however, hide a highly significant problem \times group interaction, $F(3,100) = 93.12, p < .001$, partial $\eta^2 = .737$, that is shown in Table 2. For session 1, the three worked example groups spent an average of 47 s studying the initial worked examples and 152 s solving the subsequent CT practice problems. The CT group averaged 179 s solving the CT instructional problems and 98 s solving the subsequent CT practice problems. The CT group was able to gain some time on the worked example groups on the practice problems but nonetheless required more total time because of the time required to solve the instructional problems.

An ANOVA of the learning times for the three pairs of algebra problems during the second session revealed an identical pattern. There were again significant differences in learning times for the four instructional groups, $F(3, 97) = 6.33, p < .001$, partial $\eta^2 = .164$. Bonferroni comparisons revealed that the CT group differed from each of the three worked example groups at a $p = .02$ level or lower, and the three worked example groups did not differ from each other. The CT group took more time per problem than the static table group ($M = 26.52, SD = 8.96$), the static graphics group ($M = 37.76, SD = 9.05$), and the interactive graphics group ($M = 28.17, SD = 8.88$).

The problem \times group interaction, $F(3,97) = 90.19, p < .001$, partial $\eta^2 = .736$, was again significant, as revealed in Table 2. The three worked example groups spent an average of 37 s studying the worked examples and 160 s solving the subsequent CT practice problems. The CT group spent 169 s solving the CT instructional problems and 90 s solving the CT practice problems.

A caveat is that the study times did not include the time to study the first instructional problem. It is possible that the first problem may have been more challenging in some instructional conditions than in others. For instance, the initial demands posed by an unfamiliar interactive approach (the IG condition) could be more time consuming than for a familiar interactive approach (the CT) or for a noninteractive approach (the ST and SG conditions).

CT practice problems (sessions 1 and 2)

Our next comparison of the four instructional methods examined the number of errors and the number of requested hints on the four CT practice problems in sessions 1 and 2. Figure 4a shows the number of errors made by the instructional groups on the Session-1 arithmetic problems. There was a significant effect of instruction, $F(3,105) = 10.05, p < .001$, partial $\eta^2 = .223$, in which Bonferroni post hoc tests revealed that the CT instruction was more effective than the worked examples, which did not differ among the three groups. The CT group made fewer errors per problem than the static table group

Table 2 Average learning time per problem (seconds) for sessions 1 and 2

Learning conditions	Session 1 times		Session 2 times	
	Instruct	Practice	Instruct	Practice
Cognitive Tutor	179 (6.5)	98 (11.3)	169 (5.3)	90 (10.7)
Static table	48 (6.2)	163 (10.9)	40 (5.0)	166 (10.1)
Static graphics	49 (6.5)	154 (11.3)	32 (5.1)	152 (10.3)
Interactive graphics	44 (6.2)	139 (10.9)	40 (4.9)	163 (9.9)

($M = 3.21$, $SD = .74$), the static graphics group ($M = 3.79$, $SD = .76$), and the interactive graphics group ($M = 2.81$, $SD = .75$).

There was also a significant effect of problem, $F(3, 315) = 23.18$, $p < .001$, partial $\eta^2 = .181$, and a significant problem \times group interaction, $F(9, 315) = 3.24$, $p = .001$, partial $\eta^2 = .085$. Figure 4a reveals that all the worked-example groups performed relatively well on problems 2 and 4 (which are isomorphic versions of problems 1 and 3). The findings reveal that practice on the credit card problems helped these students solve the isomorphic ore problems while students in the CT condition performed well on all problems.

The pattern of errors across the four algebra problems during the second session is almost identical (Fig. 4b). These data are based on the same 117 students from the first

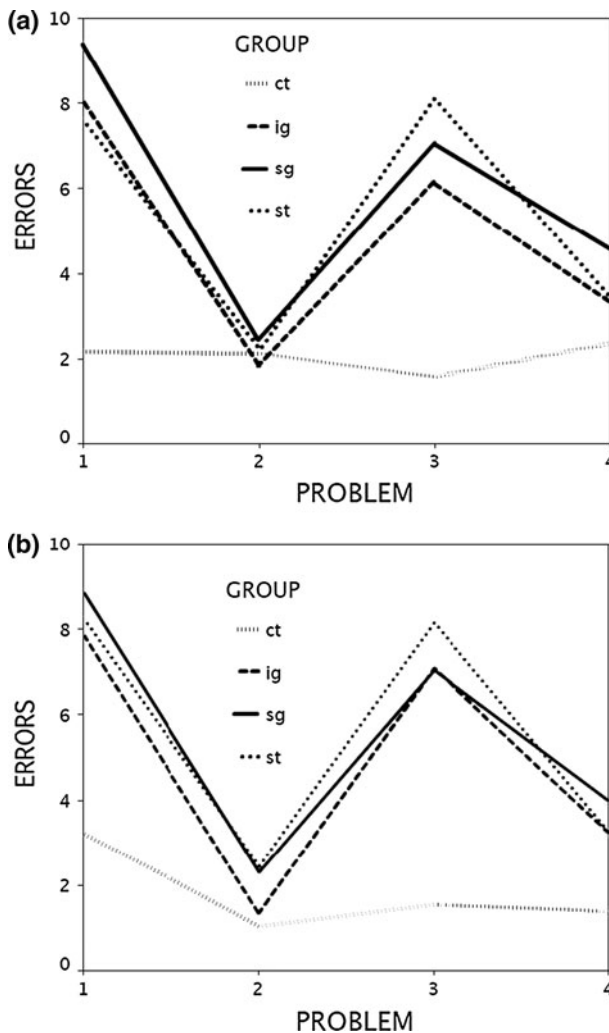


Fig. 4 Effect of instruction on CT errors on the four (a) arithmetic and (b) algebra practice problems

session except one student who made 259 errors on one problem was eliminated as an outlier. There was a significant effect of instruction, $F(3,104) = 5.82, p = .001$, partial $\eta^2 = .144$, in which Bonferroni post hoc tests revealed that the CT instruction was more effective than the worked examples, which did not differ among the three groups. The CT group made fewer errors per problem than the static table group ($M = 4.18, SD = 1.02$), the static graphics group ($M = 3.79, SD = 1.04$), and the interactive graphics group ($M = 3.16, SD = 1.03$).

There was also a significant effect of problem, $F(3, 312) = 37.39, p < .001$, partial $\eta^2 = .264$, and a significant problem \times group interaction, $F(9, 312) = 2.31, p < .05$, partial $\eta^2 = .063$. As occurred for the arithmetic word problems, students who received worked examples reduced their errors on the isomorphic ore problems. Students in the CT group made few errors on all problems.

Students could request hints from the CT as they filled out the table and constructed the equation so it is necessary to check whether fewer errors were caused by an increase in requests for hints. Figure 5a shows that students who received CT instruction requested fewer hints than the students who received worked examples, $F(3,105) = 10.66, p < .001$, partial $\eta^2 = .233$. There was also a significant effect of problem, $F(3, 315) = 70.32, p < .001$, partial $\eta^2 = .401$, and a significant problem \times group interaction, $F(9, 315) = 4.15, p < .001$, partial $\eta^2 = .106$. The interaction in this case is primarily caused by the steep drop in hint requests between the first and second problems (Fig. 5a) for students who studied the worked examples. A comparison of Figs. 4a and 5a reveals that fewer errors were typically accompanied by fewer requests for hints. The one exception is that students did not reduce their requests for hints on the final problem although they did reduce their errors.

Figure 5b reveals that students in the CT group also requested fewer hints for the algebra problems, $F(3,105) = 4.18, p < .01$, partial $\eta^2 = .107$. There was also a significant effect of problem, $F(3, 315) = 32.73, p < .001$, partial $\eta^2 = .238$, but the problem \times group interaction was not significant, $F(9, 315) = 1.89, p > .05$, partial $\eta^2 = .050$. Figure 5b reveals that all groups requested more hints for the third problem in which the unknown variable occurred twice in the equation.

In summary, solving practice problems on the Algebra CT was facilitated by also solving the preceding instructional problems on the Algebra CT. Students who received worked examples made more errors and requested more hints than students who solved all problems on the CT. These findings generalized across the three high schools; there were no significant main effects or interactions with the schools for the data displayed in Figs. 4a, b, 5a and b.

A central construct for designing instruction is transfer-appropriate processing. According to this construct, instructional task should closely duplicate the test task so students learn the information that appears on the test. For example, presenting information in a problem-oriented, rather than a fact-oriented, manner helped students use that information to solve riddles (Adams et al. 1988). Transfer-appropriate processing therefore predicts that solving the initial instruction problems on the CT should help students solve the corresponding CT practice problems on the CT.

These findings are consistent with transfer-appropriate processing in which solving problems on the CT helps students solve subsequent problems on the CT. It is unclear, however, whether the advantage of the CT over worked examples would generalize to constructing algebra equations on a paper-and-pencil posttest given at the end of the second session.

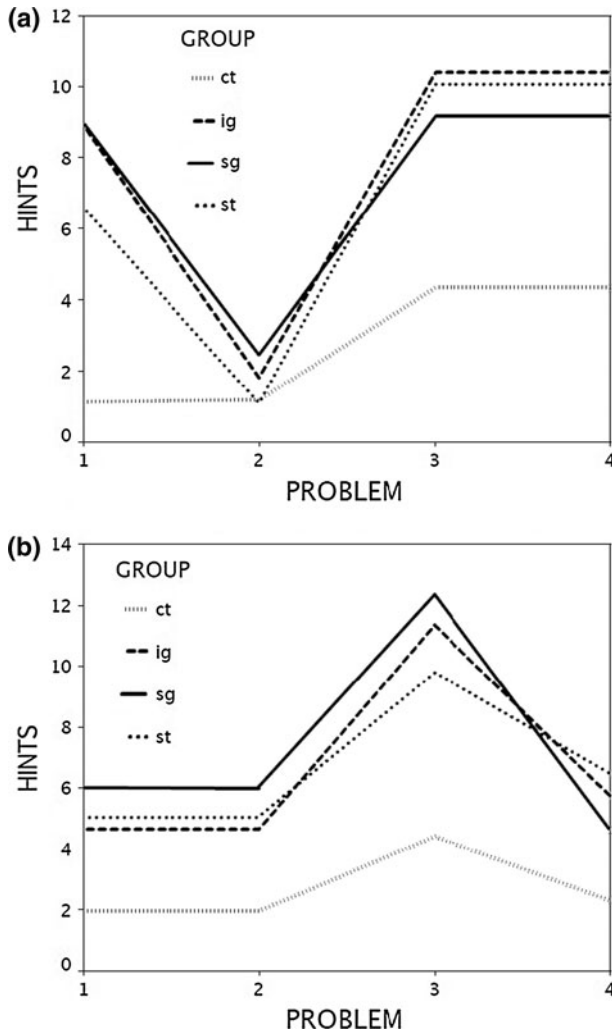


Fig. 5 Effect of instruction on CT hints on the four (a) arithmetic and (b) algebra practice problems

Paper-and-pencil pre- and posttest (session 2)

The superior performance of the CT group on the practice problems did not generalize to the paper-and-pencil posttest problems. Accuracy in constructing equations increased from 7% on the pretest to 37% on the posttest for the CT group, from 7 to 28% for the static table group, from 4 to 34% for the static graphics group, and from 7 to 28% for the interactive graphics group.

We analyzed these results in an ANOVA in which test (pretest, posttest) and problem (1, 2) were within-subject variables and instruction (static graphics, interactive graphics, static table, tutor) and school (a–c) were between-subject variables. There was a significant learning effect in which the probability of a correct response increased from .06 on the pretest to .32 on the posttest, $F(1,105) = 52.14$, $p < .001$, partial $\eta^2 = .332$. There were

also significant differences between the two problems, $F(1, 105) = 8.35, p < .01$, partial $\eta^2 = .074$ and the three schools, $F(2, 105) = 3.55, p < .05$, partial $\eta^2 = .063$. The effect of instruction was not *significant*, $F(3, 105) < 1, p > .05$, partial $\eta^2 = .012$, nor was the instruction \times test interaction, $F(3, 105) < 1, p > .05$, partial $\eta^2 = .017$. There was an interaction of school with instruction, $F(6, 105) = 2.60, p < .05$, partial $\eta^2 = .129$, but the school \times instruction \times test interaction was not significant, $F(6, 105) < 1, p > .05$, partial $\eta^2 = .042$. The four instructional conditions therefore produced similar gains from pretest to posttest across the three schools.

In summary, although the CT instruction was more helpful than worked examples for solving CT practice problems on the tutor during the learning session, the four instructional groups experienced comparable gains on the paper-and-pencil posttest.

Delayed test, transfer, and model analysis (session 3)

Results for the three tests of robust learning in session 3, delayed recall, transfer, and conceptual understanding, are displayed in Table 3. Fifteen students (two CT, five static table, four static graphics, four interactive graphics) who participated in the initial two instructional sessions were unavailable for the third session so these data are based on 102 students. Students used paper and pencil to construct equations for six word problems representing four problems that were equivalent to instructional problems and two transfer problems. Students performed better on the training than on the transfer problems, $F(1, 90) = 21.81, p < .001$, partial $\eta^2 = .195$. However, there were no significant differences among the four instructional groups, $F(3, 90) < 1, p > .05$, partial $\eta^2 = .006$, on the training and transfer problems.

The paper-and-pencil robust learning evaluation was followed by the CT Model Analysis task described in Appendix 2 in which students were asked to identify the meaning of the numbers, symbols, and terms in the equations. The model analysis required identifying a total of 31 equation parts across the four problems. The number of correctly identified parts on the first attempt was 16.36 ($SD = 6.56$) for the static graphics group, 16.13 ($SD = 7.36$) for the interactive graphics group, 18.07 ($SD = 4.54$) for the static table group, and 16.20 ($SD = 5.49$) for the CT group. An ANOVA revealed that there were no differences among the four groups, $F(3, 97) < 1, p > .05$.

Although the ANOVA revealed no significant differences, we noticed that the static table group performed consistently well in identifying components, showing the highest accuracy on 18 of the 31 components (compared to five for each of the graphics groups and three for the CT). Insignificant parametric analyses can be caused by high within-group variance as revealed by the high SDs. We therefore performed a nonparametric analysis by rank ordering the four groups for each of the 31 components in order to conduct a Friedman two-way ANOVA across components (Siegel 1956). This analysis was significant, $\chi^2(3) = 20.00, p < .001$. The sum of the ranks was 51.5 for the static table group, 77

Table 3 Robust learning measures: retention, transfer, and model analysis accuracy

Learning conditions	Retention %correct	Transfer % correct	Model analysis % correct
Cognitive Tutor	32	15	52
Static table	26	13	58
Static graphics	29	21	53
Interactive graphics	29	18	52

for static graphics, 89 for interactive graphics, and 92.5 for the CT. The range of the sums could vary from 31 (if a group always scored the highest across each of the 31 comparisons) to 124 (if a group always scored the lowest across each of the 31 comparisons). Although the parametric ANOVA across participants revealed a nonsignificant difference, a nonparametric ANOVA across items revealed instructional differences in the ability to correctly identify the components of the equations.

Discussion

Cognitive Tutor versus worked examples

The goal of this study was to evaluate four instructional conditions distinguished by three different kinds of worked example solutions (static graphics, interactive graphics, static table) and a tutor that guided students through the same arithmetic and algebra modeling problems. We proposed that these instructional approaches differ in their relative strengths of providing declarative knowledge, organizing information in a table, organizing information in a diagram, encouraging interactive learning, and creating multiple memory codes.

In order to evaluate these different strengths, it is necessary to use a variety of methods. Our evaluations included (1) the time spent on instructional and practice problems, (2) the number of errors and requested hints while using the CT to construct equations for practice problems, (3) improvement on a paper-and-pencil pre-posttest, (4) the number of correct equations on a delayed test that measured both learning and transfer, and (5) the number of correctly identified components of these equations.

The analyses of student performance on the sixteen instructional and practice problems yielded two principal results. First, we found instructional effects on constructing equations for the CT practice problems. Students who used the CT to construct equations for the instructional problems, rather than study the solutions, made fewer errors and requested fewer hints in constructing equations for the CT practice problems. This finding is consistent with transfer-appropriate processing, in which increasing the similarity between the instruction and practice situations enhances practice performance (Adams et al. 1988), and with Chi's (2009) hypothesis that interactive processing is more effective than passive processing. All students had been using the CT in lab sessions, and the result persisted through the final practice problem in each session, so the finding is unlikely caused by general unfamiliarity with using the CT. Rather, it was likely caused by providing hints and correcting errors on the preceding equivalent instructional problem. This facilitation effect, however, was achieved at a cost. The second main result for the learning sessions is that students took longer to solve the instructional problems on the CT so took more time to complete the learning sessions.

However, the greater time on task in learning for the CT group and the greater success of the CT group in solving the CT practice problems did not translate into greater posttest problem-solving accuracy. Instead, there were no reliable differences among the four groups on pretest–posttest learning gain in session 2. This finding is consistent with previous studies in other domains (chemistry, geometry) that found that incorporating worked examples into tutor-supported problem solving can increase learning efficiency—reducing the learning time to achieve similar learning gains on problem solving tests.

On the whole, there were also no differences among the four groups on the robust learning tasks designed to measure retention of problem-solving knowledge and transfer to novel problems. The only difference among the four learning conditions emerged in the

model analysis task in which students who received the static table worked examples were more successful than the students in any of the other three conditions in identifying the meaning of the hierarchical equation components. Among the three worked-example conditions, this finding is consistent with transfer-appropriate processing because the column and row labels in the tables (Fig. 1) associate verbal descriptions with the equation components, while the graphic representations do not.

The difference between the static-table worked-example and CT problem-solving condition for this robust learning task is particularly interesting, because both of these representations employ the same table format with the same verbal labels, but in contrast with Chi's (2009) hypothesis and case studies, the relatively passive worked examples are more successful than interactive CT problem solving. This appears to be a case in which transfer-appropriate processing trumps interactive processing. The chief activity in the static table worked examples is to read further verbal descriptions of how to represent components of the problem situation in a mathematical model, while the chief activity in CT problem solving is to actually construct the model. The relatively passive verbal activity appears to transfer more readily to the model analysis task than does the interactive equation construction activity.

The immediate problem-solving test and transfer test results are consistent with this transfer-appropriate analysis. There were no differences between static table condition and CT conditions on these two tests. Students in both conditions had relevant experience in writing mathematical equations during learning, in the CT practice problems, and that knowledge transferred equivalently to the two types of model generation tests. The benefit of static table worked examples only emerged when students were asked to perform the verbal model analysis task.

It should also be noted that the learning outcomes were modest for all instructional conditions, with students averaging 32% correct on the immediate problem-solving post-test during the second session. These were challenging problems for the students, based on their pretest scores and a likely cause of the modest outcomes is that students received only two training sessions. That students achieved 25 percentage-point learning gains after devoting 12–20 min to each problem type is promising and more time on task would seem warranted. Another likely cause is that students did not participate in more conceptually oriented supplemental activities such as providing self explanations (Berthold et al. 2009).

Variations in worked examples

This study strongly replicates earlier findings in mathematics and chemistry that incorporating worked examples with tutor-supported problem solving can improve learning efficiency. A difference between our study and previous studies that combined the CT with worked examples is that we compared worked examples that either incorporated a static table, static graphics, or interactive graphics. What can be concluded about which forms of worked examples to include? As indicated in Table 1 these worked examples differ along a number of dimensions. Differences in representations can have both advantages and disadvantages for learning (McNeil et al. 2009). Ainsworth (2006) has therefore proposed that an advantage of using multiple representations is that they can complement each other by combining different strengths.

A difference between research and practice is that different instructional approaches, such as the three variations of worked examples, are typically competitors in research but can become collaborators in practice. For example, Reed (2005) discusses how he compared the relative effectiveness of instruction based on either simulations, computations, or

conceptual reasoning in research (Reed and Saavedra 1986) but integrated the three types of instruction in designing the Animation Tutor: Average Speed module.

Reed and Hoffman (2011) also used an integrated instructional approach in designing the Animation Tutor: Mixtures module that can be downloaded from the Animation Tutor web site <http://www.sci.sdsu.edu/mathtutor/index.htm>. The module combines the complementary strengths of interactive graphics (interactive learning, multiple memories) and static tables (mapping concepts to mathematical symbols). The worked examples in the Mixture module consist of the interactive graphics instruction (Fig. 2) for the algebra problems in Appendix 1. Each of the worked examples is followed by an equivalent test problem that requires the computation of a numerical answer. Each test problem is then followed by a detailed solution based on the interactive table instruction (Fig. 1).

The interactive graphics condition might have been more effective if students had to explain why manipulating one bar does or does not affect other bars. Interventions that encourage self explanations have enhanced the effectiveness of worked examples (Atkinson and Renkl 2007; van Gog et al. 2004), including the integration of multiple representations (Berthold et al. 2009). A challenge, however, in comparing different instructional methods is that self-explanations would differ across methods. We therefore decided that initially establishing a “baseline” comparison without differential scaffolding would be more valuable in our initial research.

However, withholding scaffolding is not necessary in practice. The interactive graphics instruction in the Animation Tutor: Mixture module therefore requires that students explain why adjusting one bar influences other bars to encourage them to reflect conceptually on their manipulations. After receiving the worked examples for the first two algebra problems in Appendix 1, students type in an answer to the questions:

You were able to vary the variable by raising and lowering one of the bars. What other bar changed as you raised and lowered this bar? Explain how and why this bar changed?

Similar questions followed the last two worked examples in Appendix 1 in which varying the variable changed the height of both of the other bars. The instruction provided detailed explanations following students’ self-explanations.

Although interweaving interactive graphics with static table instruction and eliciting conceptual explanations are promising additions to the worked-example instruction, there is a need for continued interweaving of instructional design with research to assess instructional modifications (Reed 2005). Prompting self-explanations is not always effective (Gerjets et al. 2004). Interactions between design parameters, the functions of multiple representations and the cognitive tasks (Ainsworth 2006) should keep instructional researchers and designers busy in enhancing and applying knowledge.

Appendix 1: Worked example problems

Session 1

Arithmetic type 1 interest

1. You have a MasterCard with a balance of \$532 at a 21% interest rate. You also have a Visa credit card with a balance of \$841 at a 16% interest rate. How much money are you paying in total annual interest?

$$.21 \times \$532 + .16 \times \$841 = \text{Total Interest}$$

Arithmetic type 1 ore

2. The Acme Mine yielded 67 tons of low-grade (35% iron) ore in the first year of operation. In the second year of operation, they yielded 62 tons of high-grade (75% iron) ore. How much pure iron did the company mine in two years?

$$.35 \times 67 \text{ tons} + .75 \times 62 \text{ tons} = \text{Total Iron}$$

Arithmetic type 2 interest

3. At the end of her first year of using credit cards, Shelly owed \$475 in total interest on her MasterCard and Visa accounts. Her MasterCard charges 19% interest and her Visa Card charges 22% interest. She paid the interest on her Visa Card debt of \$1100. How much interest does she still owe on her MasterCard?

$$\$475 - .22 \times \$1100 = \text{Remaining Interest}$$

Arithmetic type 2 ore

4. A jeweler ordered 40 oz of pure gold. They supplier extracted the 40 oz from sylvanite gold ore that contains 28% pure gold and calaverite gold ore that contains 40% pure gold. If the supplier used 55 oz of calaverite ore, how much gold came from the sylvanite ore?

$$40 \text{ ounces} - .40 \times 55 \text{ ounces} = \text{Gold from Sylvanite Ore}$$

Session 2

Algebra type 1 interest

1. You have an American Express credit card with a balance of \$715 at an 11% interest rate and a Visa credit card with a 15% interest rate. If you pay a total of \$165 in annual interest, what is the balance on your Visa card?

$$.11 \times \$715 + .15 \times V = \$165$$

Algebra type 1 ore

2. The Acme Mine yielded 67 tons of low-grade (35% iron) ore. How many tons of high-grade (75% iron) ore will they need in order to yield 70 total tons of pure iron?

$$.35 \times 67 \text{ tons} + .75 \times H \text{ tons} = 70 \text{ tons}$$

Algebra type 2 interest

3. You have a total balance of \$1405 on two different credit cards— an American Express credit card with a 12% interest rate and a Discover credit card with a 24%

interest rate. If you owe a total of \$224 in annual interest, what is your balance on the Discover card?

$$.24 \times D + .12 \times (\$1405 - D) = \$224$$

Algebra type 2 ore

4. You have 51 oz of gold ore. The sylvanite gold ore contains 28% pure gold and the calaverite gold ore contains 40% pure gold. If you have 16 total ounces of pure gold, how many ounces of calaverite gold ore do you have?

$$.40 \times C \text{ ounces} + .28 \times (51 - C) \text{ ounces} = 16 \text{ ounces}$$

Appendix 2

CT Model Analysis problems

Each model analysis problem presents both a real-world situation and a symbolic model of the situation. Students are asked to describe what each hierarchical component of the symbolic model represents in the problem situation.

Connected - AutoTime - v. 2.0.00
Hint

A mine extracted 91 tons of copper ore. They mined both high-grade (81% copper) ore and low-grade (32% copper) ore. If they had 42 total tons of pure copper, how many tons of low-grade (32% copper) ore were mined? Using L as your variable you can represent this situation with the equation: $0.32L + 0.81(91 - L) = 42$

This mathematical expression	represents:
<input type="text" value="L"/>	---
<input type="text" value="0.32"/>	---
<input type="text" value="0.32L"/>	---
<input type="text" value="91"/>	---
<input type="text" value="91 - L"/>	---
<input type="text" value="0.81"/>	---
<input type="text" value="0.81 (91 - L)"/>	---
<input type="text" value="0.32L + 0.81 (91 - L)"/>	---
<input type="text" value="42"/>	---

Equation part	Answers
L	English description
$.32$	The number of tons of low-grade ore they have
$0.32L$	The percent of copper in the low-grade ore
91	The number of tons of pure copper in their low-grade ore
$(91 - L)$	The total tons of copper ore they have
$.82$	The number of tons of high-grade ore they have
$.82(91 - L)$	The percent of copper in the high-grade ore
$0.16L + .82(91 - L)$	The number of tons of copper in their high-grade ore
42	The total number of tons of pure copper they have

References

- Adams, L. T., Kasserman, J. E., Yearwood, A. A., Perfetto, G. A., Bransford, J. D., & Franks, J. J. (1988). Memory access: The effects of fact-oriented versus problem-oriented acquisition. *Memory & Cognition*, *16*, 167–175.
- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*, 183–198.
- Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education (IJAI-IED)*, *19*, 105–154.
- Anderson, J. R. (1990). Analysis of student performance with the LISP tutor. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 27–50). Hillsdale, NJ: Erlbaum.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, *4*, 167–207.
- Anderson, J. R., & Fincham, J. M. (1994). Acquisition of procedural skills from examples. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *20*, 1322–1340.
- Anthony, L. (2008). *Developing handwriting-based intelligent tutors to enhance mathematics learning*. Unpublished doctoral dissertation, Carnegie Mellon University.
- Atkinson, R. K., & Renkl, A. (2007). Interactive example-based learning environments: Using interactive elements to encourage effective processing of worked examples. *Educational Psychology Review*, *19*, 375–386.
- Barsalou, L. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, *22*, 577–660.
- Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, *37*, 345–363.
- Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology*, *101*, 70–87.
- Chi, M. T. H. (2009). Active–constructive–interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*, 73–105.
- Cooper, G., & Sweller, J. (1987). The effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, *79*, 347–362.
- 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* (pp. 245–252). New York, ACM Press.
- Corbett, A., McLaughlin, M., Scarpinato, K. C., & Hadley, W. (2000). Analyzing and generating mathematical models: An Algebra II Cognitive Tutor design study. In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Intelligent tutoring systems: 5th international conference* (pp. 314–323). New York: Springer.
- Corbett, A. T., Wagner, A., Lesgold, S., Ulrich, H., & Stevens, S. (2007). Modeling students' natural language explanations. In C. Conati, K. McCoy, & G. Paliouras (Eds.), *User Modeling 2007: Proceedings of the eleventh international conference* (pp. 117–126). Berlin: Springer.

- Corbett, A., Wagner, A., & Raspat, J. (2003). *The impact of analysing example solutions on problem solving in a pre-algebra tutor*. Proceedings of AIED 2003: The 11th international conference on artificial intelligence and education.
- Engelkamp, J. (1998). *Memory for actions*. Hove: Psychology Press.
- Fuson, K. C., & Willis, G. B. (1989). Second grader's use of schematic drawings in solving addition and subtraction problems. *Journal of Educational Psychology, 81*, 514–520.
- Gamo, S., Sander, E., & Richard, J. F. (2010). Transfer of strategy use by semantic recoding in arithmetic problem solving. *Learning and Instruction, 20*, 400–410.
- Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing instructional examples to reduce intrinsic cognitive load: Molar versus modular presentation of solution procedures. *Instructional Science, 32*, 33–58.
- Goldstone, R. L., Landy, D. H., & Son, J. Y. (2009). The education of perception. *Topics in Cognitive Science, 2*, 265–284.
- Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *The Journal of the Learning Sciences, 14*, 69–110.
- Heffernan, N. T., & Koedinger, K. R. (1998). A developmental model for algebra symbolization: the results of a difficulty factors assessment. In *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 484–489). Hillsdale, NJ: Erlbaum.
- Hmelo-Silver, C. E., Duncan, R. V., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*, 99–107.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23–31.
- Kalyuga, S., Chandler, P., & Sweller, J. (2001a). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*, 5–23.
- Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001b). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*, 579–588.
- Kirschner, P. A., Sweller, J., & Clark, R. C. (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., & Alevan, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19*, 239–264.
- 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.
- MacGregor, M., & Stacey, K. (1997). Students' understanding of algebraic notation. *Educational Studies in Mathematics, 33*, 1–19.
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge: Cambridge University Press.
- McLaren, B. M., Lim, S., & Koedinger, K. R. (2008). When and how often should worked examples be given to students? New results and a summary of the current state of research. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 2176–2181). Austin, TX: Cognitive Science Society.
- McNeil, N. M., Uttal, D. H., Jarvin, L., & Sternberg, F. J. (2009). Should you show me the money? Concrete objects both hurt and help performance on mathematics problems. *Learning and Instruction, 19*, 171–184.
- Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*, 309–326.
- Ng, S. F., & Lee, K. (2009). The model method: Singapore children's tool for representing and solving algebraic word problems. *Journal for Research in Mathematics Education, 40*, 282–313.
- Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*, 429–434.
- Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1–8.
- Paas, F. G. W. C., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*, 122–133.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. New York: Oxford University Press.
- Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

- Philipp, R. A. (1992). The many uses of algebraic variables. In B. Moses (Ed.), *Algebraic thinking, Grades K-12: Readings from NCTM's School-based journals and other publications* (pp. 157–162). Reston, VA: National Council of Teachers of Mathematics.
- Reed, S. K. (1987). A structure-mapping model for word problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 124–139.
- Reed, S. K. (1999). *Word problems: Research and curriculum reform*. New York: Erlbaum.
- Reed, S. K. (2005). From research to practice and back: The Animation Tutor project. *Educational Psychology Review*, *17*, 55–82.
- Reed, S. K. (2006). Cognitive architectures for multimedia learning. *Educational Psychologist*, *41*, 87–98.
- Reed, S. K. (2008). Manipulating multimedia materials. In R. Zheng (Ed.), *Cognitive effects of multimedia learning* (pp. 51–66). New York: IGI Global.
- Reed, S. K. (2010). *Thinking visually*. New York: Taylor & Francis.
- Reed, S. K., & Ettinger, M. (1987). Usefulness of tables for solving word problems. *Cognition and Instruction*, *4*, 43–59.
- Reed, S. K., & Hoffman, B. (2011). *Animation Tutor: Mixtures. Instructional software*. San Diego: San Diego State University.
- Reed, S. K., & Saavedra, N. C. (1986). A comparison of computation, discovery, and graph procedures for improving students' conception of average speed. *Cognition and Instruction*, *3*, 31–62.
- Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skills acquisition: A cognitive load perspective. *Educational Psychologist*, *38*, 15–22.
- Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, *14*, 249–255.
- Salden, R. J. C. M., Alevin, V., Schwonke, R., & Renkl, A. (2010). The expertise reversal effect and worked examples in tutored problem solving. *Instructional Science*, *38*, 289–307.
- Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Alevin, V., & Salden, R. (2009). The worked-example effect: Not an artifact of lousy control conditions. *Computers in Human Behavior*, *25*, 258–266.
- Schwonke, R., Renkl, A., Salden, R., & Alevin, V. (2011). Effects of different ratios of worked solution steps and problem solving opportunities on cognitive load and learning outcomes. *Computers in Human Behavior*, *27*, 58–62.
- Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 215–266). San Diego: Academic 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.
- Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, *80*, 424–436.
- 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.
- Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*, 1–39.
- Weitz, R., Salden, R. J. C. M., Kim, R. S., & Heffernan, N. T. (2010). Comparing worked examples and tutored problem solving: Pure vs. mixed approaches. In S. Ohlsson & R. Catrambone (Eds.), *Cognition in flux: Proceedings of the thirty-second annual meeting of the cognitive science society* (pp. 2876–2881). Austin TX: Cognitive Science Society.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*, 625–636.