Scaffolding Problem Solving with Annotated, Worked-Out Examples to Promote Deep Learning

Michael A. Ringenberg and Kurt VanLehn

University of Pittsburgh, Learning Research and Development Center

3939 O'Hara St., Pittsburgh PA 15260, USA

412-624-3353

mringenb@pitt.edu, vanlehn@cs.pitt.edu

http://www.pitt.edu/~mringenb/

Abstract. This study seeks to compare the relative utility for learning college-level physics of intelligent tutoring systems that have procedural based hints and worked-out examples. In order to test which produced better gains, a modified version of Andes was used in which participants either received hints or annotated, worked-out examples in response to their help requests. We found that providing annotated, worked-out examples instead of hint sequences was more efficient in the number of problems it took to obtain basic mastery.

1 Introduction

At the heart of all educational research is the search for ways to improve the instruction of novices. One strategy that has been found to be very effective is one-on-one human tutoring [1]. The economics of providing one-on-one tutoring has prompted the investigation of other techniques to aid instruction. One technique has been the development of intelligent tutoring systems, which are computer programs that supplement classroom instruction in order to promote learning as if the student is receiving individual instruction. Another technique is using examples in instructional material to increase its effectiveness [2], [3], [4], [5], [6]. As both paths have meet with some success, it is worth exploring ways to combine them.

Our exploration was done by modifying Andes, an intelligent tutoring system that aids the instruction of college-level introductory physics [7]. Andes's main function is to present a student with a problem and to let the student solve it while receiving adaptive scaffolding from the system. See Figure 1 for a screen shot of the interface.

The scaffolding has two major forms. One is flag feedback and the other is hints. The flag feedback is the marking of the student's input as either correct or incorrect. When the student asks for help, Andes tries presenting the student with a hint. The hints either point out what is wrong or suggests a step to do next. These hints are based on identifying what steps that the student has completed in a solution that is generated from basic definitions and principles of physics. Because of this approach, the hints tend to address and point to basic

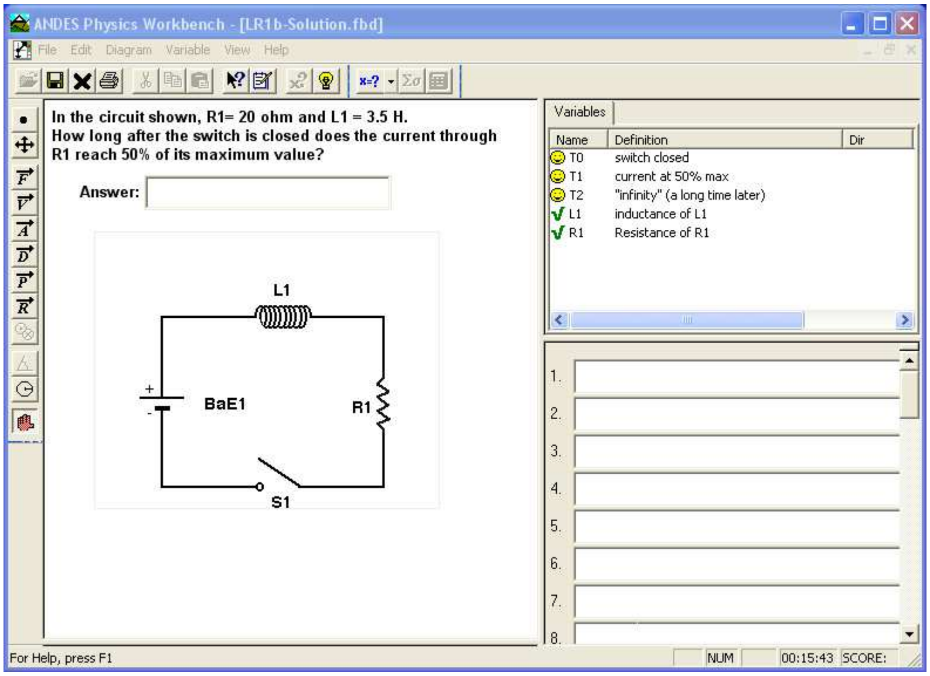


Fig. 1. The Andes Physics Workbench, an intelligent tutoring system for helping student solve physics problems. The problem shown is from the Inductors section of problems.

principles and definitions. This is desirable because it tries to link the current problem and step with what the student already knows.

The hints are also presented in a sequential fashion known as a hint sequence. The student is typically presented with a very vague hint first in the hope that it will jog the students memory or prompt the student to self explain the next step or the current error. The student can then ask for the next step in the hint sequence. The hints become more concrete as the sequence is followed. The last hint in a sequence is typically the correct step that the student should enter.

The last hint is referred to as the bottom-out hint. This structure of hints has been used in several Intelligent Tutoring Systems and an example of the hints provided by Andes can be seen in figure 2.

Students can and do abuse this form of help [8]. Students will click through the hints to in order to get the bottom-out hint quickly and ignore the rest of the hint sequence. This help abuse is a problem because it is associated with shallow learning [8].

Our basic hypothesis is that the learning of novice students will increase if we replace Andes's hint sequences with worked-out examples. A worked-out example is a problem with all of the necessary steps needed to solve it. There are several reasons to believe that worked-out examples will be more effective than hint sequences for novices.

First, from anecdotal evidence of previous Andes studies, some students will ask for the bottom-out hints for most of the steps in a problem and then refer back to that problem when they attempt to solve a similar problem. In essence, they are creating a worked-out example. This anecdotal evidence is consistent with studies showing that novices prefer to learn from the examples as opposed to written, procedural instructions [9].

Second, we suspect that the hints provided by Andes can provide good targeted help to students who are already familiar with the domain and have an adequate understanding of the underlying principles. However, for novices in the domain, the first couple of hints in the hint sequence probably make little sense as the novices are not sufficiently familiar with the principles for the hints to activate the proper reasoning to finish the given step nor are they familiar enough with the problem structures to understand why the hint is relevant.

Third, worked-out examples can be effective instruction. In one study, worked-out examples were more effective than presenting procedural rules [3]. However, examples are more effective when they alternate with problem solving, presumably because studying large blocks of examples becomes boring [10]. By using a single example in place of a hint sequence, we can avoid the boredom of example blocks.

On the other hand, worked-out examples are not always effective. There effectiveness seems to require that students self-explain them. A self-explanation for an example is a student utterance that is a meaningful and correct explanation of a step [11]. Unfortunately, students tend not to spontaneously produce self-explanations and poor students tend to produce ineffective self-explanations. Useful self-explanations tend to be either derivations or procedural explanations [12]. Derivations basically answer the question: “Where did this step come from?” Procedural explanations answer the question: “Why was this step done?”

If this process is not done, the student is not likely to have a deep understanding of the material. It is exemplified by the observation that students who use examples poorly use surface feature of the problem and the example as criteria for matching whereas experts will use the principles and deep structure as the criteria [13].

By providing worked-out examples with annotations, we hope to encourage students to learn the problem structures and see examples of good self-explanations. The annotations provided are the principles used to derive each step and the solution is presented in a well structured order. The annotated, worked-out examples should provide a context in which to situate the principle application. Also, providing the annotated, worked-out example during problem solving should encourage the student to focus on the common deep structure between the problem and the example. This will lead to students who are provided with annotated, worked-out examples to perform better on tasks that test the deep structural understanding of problems in the domain.

2 Methodology

2.1 Overview

This study was a hybrid study in that it was both naturalistic and experimental. The experiment was conducted during the inductors section of a second semester, college-level, physics course. As part of the course, students solved homework problems with Andes. Students who volunteered to participate in the experiment used a modified version of Andes to do this homework. However, they did their homework when and how they pleased, just as they always did. For instance, some students did this homework right before the in-class exam, which was days after the post-test.

The study had two conditions: the examples condition and the hints condition. In the examples condition participants were presented with annotated, worked-out examples in response to their help requests while using Andes. In the hints condition, participants were given Andes's normal hints in response to help requests. The dependent measure for this experiment was a problem matching task where participants were asked to pick which problem statement from a choice of two would be solved most similarly to a given problem statement.

2.2 Population

The participants in this study were recruited from the Pittsburgh Science of Learning Center LearnLab (www.learnlab.org). At the time of this study, the Physics section of the LearnLab was run as part of the General Physics I/II classes in the United States Naval Academy in Annapolis, Maryland. A total of forty-six participants were recruited from two sections of this course. Both sections were taught by the same professor.

2.3 Environment

Participants were instructed to solve their homework using a modified version of Andes. The participants were already familiar with Andes as it has been used throughout the course as part of the assigned homework. All students in the LearnLab used Andes on their personal computers or in computer labs.

Participants were not restricted in any way as to their access to textbooks, professors, peers, or any other supplementary material during the training part of the experiment. They used the Andes system on their own time and at their own pace.

The modification to Andes for the study was to add near transfer worked-out examples for the inductance problems. When a participant in the examples condition asked for help, they were presented with the example that corresponded to the inductor problem they were working on instead of Andes's normal hint sequence. In all other situations, the modified Andes acted as it normally would.

The post-test was conducted during the lab session prior to the in-class examination on the material. The assessment was a computer based, multiple choice test with twenty questions presented in random order. The participants were given thirty minutes to complete this assessment. For the exact instructions given to the participants and a sample question from the post-test, see Appendix A.3.

2.4 Materials

There were two major components of this experiment that had to be created, the examples and the post-test.

The worked-out examples were designed to be near-transfer problems to ones that the participants were solving. Near-transfer problems were defined to be problems that have the same deep structure; in other words, ones that used the same principles in their solutions. Numeric values were changed as well as some other minor surface features changes. Often the biggest change between a training problem and its corresponding example was which quantity was sought. For example in problem A.1, current, time, and the inductance were given and the sought quantity was the voltage. In the corresponding example, A.2: the inductance, the voltage, and the current were given and the time was the sought quantity. Both problems use the principles of the voltage across an inductor and the definition of the constant rate of change.

Problems that were solved using the same principles shared a common worked-out example. This led to an average of two problems paired with each of the five worked-out example.

The example problems were solved by an expert, who then went through each solution and annotated the solution with the principle used in each step or listed the equations algebraically combined for a given step. These annotations were listed next to each step of the worked-out example and were visible to the participant. The format of the examples was designed to resemble the solutions a participant would generate in the Andes environment; see Appendix A.2 to view one of the worked-out examples. The names of the principles used in the worked-out examples were the same names that Andes used in its hints and in its incorporated textbook pages. The principles in the examples were linked to the appropriate Andes textbook page. This was done so that the information available to the participants in the examples and hints conditions was the same.

The post-test was adapted from the similarity judgment task of [14]. Given a model problem, participants had to choose which of two comparison problem were solved most similarly to the model problem. All three problems were unsolved and the participants only saw a problem statement comprised of a few sentences and a diagram, see Appendix A.3. Each of the comparison problem could be classified into one of four relationships to the model problem. (1) They could have the same surface features which means that the problems were similar in format and the same diagram was used, but they were solved differently. (2) They could have the same deep structure which means that the surface features were different but the problems were solved using the same principles. (3) They could have the same surface features and deep structure. (4) Finally, they could differ in both surface features and deep structure.

A general principle for constructing the questions in the similarity judgment task was that only one of the comparison problems matched the model problem statement with respect to deep structure. This limits the number of possible combinations of types to four different comparison pairs, either surface/deep, surface/surface&deep, none/deep, and none/surface&deep. In the target domain of this study, there were five different problem structures practiced. This produced a total of twenty question, one of each comparison pair type fore each structure.

One important difference between what [14] did and what was done in this study was that the model problem was not necessarily the same for each of the five different problem structures and that problems used a model problems were reused as comparison problems.

The post-test questions were comprised of problems from the training set, the worked-out examples, and other generated problems. These questions were presented to each participant in a random order and the comparison pair items were ordered randomly.

2.5 Methods

The participants were assigned to one of two conditions, either the examples condition or the hints condition. The participants were assigned in a pairwise random fashion based on their cumulative GPA and irrespective of any other consideration such as section, gender, or age, in order to balance the two conditions in regards to previous performance.

Participants were instructed to solve their assigned homework using the modified version of Andes. The participants were on their own recognizance to do the assigned work. Ten of the fourteen inductors problems in Andes were assigned as part of the homework for the inductors section of the class.

In the lab session prior to the in-class examination on the material, the participants were given the post-test. They had thirty minutes to complete it.

3 Results

When the Andes log files were examined, it was discovered that eighteen of the forty-six participants had not solved any physics problems involving inductors with Andes before the post-test was administered. Nine participants solved problems with examples as aids. Twelve participants solved problems with hints. Five participants solved problems but refused any help. Finally, two participants were dropped from the analysis due to their failure to complete the post-test or any of the Andes problems.

Of the eighteen participants who did not solve any problems before the post-test, five of them solved some of the problems after the post-test was administered and before the exam covering the same material. An additional seven participants from this group solved problems after the exam and before the final exam. There is also evidence that participants from the other conditions also solved additional problems and reviewed solved problems after the post-test and before the exam.

The five participants who refused any help were dropped from the analysis for several reasons. One reason was that they do not fit into any of the other groups as they solved problems but refused help. Refusing help was good evidence that these participants did not think that the help provided by Andes was useful in the past and did not think it would help them with the problems in this study. It was also discovered that two of these participants were already experts in the material. Additionally, all of them were statistical outliers.

Though the study was designed to have only two conditions, the examples condition and the hints condition, it was decided to add a third condition for the eighteen participants who did not complete any problem solving before the post-test instead of dropping then from all of the analyzes.

There was no significant difference in performance on circuit questions between the three conditions on the in-class examination administered before this study ($\mu_{no\ ps} = 187.2$ $\sigma_{no\ ps} = 42.5$; $\mu_{hints} = 178.4$ $\sigma_{hints} = 53.3$; $\mu_{examples} = 200.9$ $\sigma_{examples} = 49.6$; $F_{(2,36)} = 0.57$ $p = 0.5684$). This suggests that even though the participants selected whether they were in the no-problems-solved condition or one of the other two, the conditions ended up with equivalently competent students.

Next, we discus training time. There was a large significant difference between the no-problems-solved condition and each of the other two, in other words, both other conditions worked on problems for an amount of time significantly greater than zero ($p < 0.00001$). There was a large, but not significant difference in the total amount of time spent solving problems with the hints group spending more time solving problems than the examples group ($\mu_{hints} = 7942$s $\sigma_{hints} = 5793$; $\mu_{examples} = 4189$s $\sigma_{examples} = 3132$; $t_{17.6\ corrected} = 1.90$ $p = 0.0735$). However the average time per problem ($\mu_{hints} = 672$s $\sigma_{hints} = 375$; $\mu_{examples} = 508$s $\sigma_{examples} = 210$; $t_{19} = 1.18$ $p = 0.2540$) was not significantly different.

There was a significant difference in the average number of problems solved by the participants in the two problem solving conditions ($\mu_{hints} = 11$ $\sigma_{hints} = 2.4$; $\mu_{examples} = 8.1$ $\sigma_{examples} = 3.7$; $t_{19} = 2.17$ $p = 0.0427$) with the examples group solving fewer problems on average.

The participants were allowed up to thirty minutes to complete the post-test. No participant took longer than twenty-two minutes to complete the task, with twenty-eight of the participants completing it in less than ten minutes. There was no significant difference on time to finish the post-test between the three conditions ($\mu_{no\ ps} = 474$s $\sigma_{no\ ps} = 192$; $\mu_{hints} = 652$s $\sigma_{hints} = 281$; $\mu_{examples} = 657$s $\sigma_{examples} = 312$; $F_{(2,36)} = 2.47$ $p = 0.0986$).

By construction, the post-test problems varied considerably in their difficulty. For instance the none/surface&deep problems should be much easier than surface/deep problems. To more accurately measure competence, a weighted score was used. The post-test problems were weighted according to their difficulty as determined by the performance of the no-problems-solved participants on each of the problems. The weight for each problem was $1 - \frac{\#correct}{18}$, where $\#correct$ was the number of participants who correctly answered the given question from the no-problems-solved group. The weights on the problems were checked against expectations of performance on the different problem types. The weights did correspond with expectations such that participants did well on questions in the category of none/surface&deep, did poorly on surface/deep questions, had average performance on surface/surface&deep, and did moderately well on none/deep questions.

When the ANOVA model was fit using the weighted post-test score, a difference between the three groups with regards to the weighted post-test score was detected ($F_{(2,36)} = 8.49$ $p = 0.0010$). With the Tukey-Kramer adjustment for multiple comparisons, it was found that the participants in the examples condition did significantly better on the post-test than those in the no-problems-solved group ($t = 3.98$ $p = 0.0009$). The hints group also did better than the no-problems-solved group ($t = 2.45$ $p = 0.0496$). However, it was not possible to distinguish a difference between the hints group and the examples group based solely on the weighted post-test score ($t = 1.61$ $p = 0.2525$), see figure 3.

We also measured the efficiency of training, that is, the amount of gain per training problem. Because the examples and hints conditions had the equivalent weighted post-test scores and the the examples condition used fewer training problems, the efficiency (weighted post-test score/problems) was better for the examples condition ($t = 3.34$ $p = 0.0034$), see figure 4.

Next, we checked for non-linearity in efficiency. As discussed earlier, we hypothesis that examples have advantages over hints during the first few exposures, but as students become experts, examples loose this comparative advantage.

Thus, efficiency should be non-linear. An ANCOVA model was specified with the weighted post-test score as the dependent variable and the independent categorical variable was whether the participant received hints or examples during their problem solving and the number of problems completed was added as a covariate. It was determined that the slope of the regression lines for the two groups were equal, in other words, there was no significant interaction effect ($p = 0.8290$), see figure 5. When the number of problems was set to the overall mean ($\mu_* = 9.76$), the difference in the means of the weighted post-test score was significant ($t = 2.30$ $p = 0.0338$) with the examples group mean weighted score ($\mu_{examples} = 4.88$ $\sigma^2_{examples} = 0.24$) being higher than the hints group's score ($\mu_{hints} = 4.14$ $\sigma^2_{hints} = 0.22$) which is consistent with the t-test above concerning the problem efficiency. Because the slopes are similar over the range of three to fourteen problems solved by the students, yet the examples line was higher, the first one to three examples must have provided most of the comparative advantage of examples over hints.

4 Discussion

4.1 Conclusions

The problem matching task used as the post-test in this study has been used in the past to assess competence and is argued to be good at assessing expertise because it is sensitive to the knowledge of the deep structure of the problems and did not rely on algebraic competence [14]. From the results of the post-test using all of the conditions, it appears to show that studying and solving problems does improve performance in this task as both the hints and examples groups did significantly better than the no-problems-solved group.

Since there was a self selection concern, the performance on the in-class examination covering basic circuit concepts, Kirchoff's Law, capacitors, and Ohm's Law, was analyzed and it was found that there was no discernible difference in prior knowledge of circuit concepts. This implied that the conditions were balanced in regards to prerequisite knowledge of circuits.

Another concern was the difference in the amount of time the participants spent problem solving, in other words, there was a concern about time on task differences. There was a difference in the number of problems solved before the post-test. The participants who received examples solved fewer problems on average than the participants who received hints. This was not due to the participants in the examples condition taking longer to solve problems, as the average time to solve a problem was not significantly different. The analysis of the time spent solving problems can suggest that participants in the hints condition may have spent more time solving problems than the participants in the examples condition. This was easily explained by the evidence that the participants in the hints condition solved more problems than the participants in the examples condition.

Even though participants in the examples condition solved fewer problems than the participants in the hints condition, they did just as well as the participants in the hints condition on the post-test. Thus, the examples were more efficient overall than the hints. The analysis of the hints condition versus the examples condition with the number of problems as a covariate showed that after the first few problems, the gain per problem was the same in the two conditions this suggests that the examples were more efficient than the hints only for the first few problems.

There is the concern that studying worked-out examples can lead to shallow learning [13], but this study seems to indicate that worked-out examples may in fact lead to deeper learning as indicated by the post-test scores of the examples group being higher than the no-problems-solved condition and equivalent to the hints condition with fewer problems solved.

These results do suggest that examples were certainly no worse than hints in this domain and in fact were more efficient. This has implications for tutoring system design. It suggests that replacing hints with examples can be an important addition to Intelligent Tutoring Systems especially for the first few problems of each problem set. Moreover, they can be fairly cheap in terms of implementation time to add to an Intelligent Tutoring Systems, so replacing *all* hints with examples is attractive and should not hurt learning. The examples used in this study were easily added to Andes, mostly because there was no "intelligence" that needed to be added to insert examples into the system, just a simple mapping of problem names to their relevant examples.

One possible shortfall of hint sequences is that they are typically too short to effectively establish a shared context with the participant. If the intelligent tutoring system that provides hints are very good at assessing the participant's thought process, then the hints it provides are probably very effective when the system and the participant are on the same page. However, if there is any mismatch, then the hints are probably close to meaningless to the participant and if this happens too much, the participant will probably evaluate the systems hints as useless. Examples can be a way of making sure the participant and the system are on the same page by having a concrete shared context. Examples could be used when the system's student model has not collected enough data to make good evaluation of the participant or if communication seems to have failed.

4.2 Future Work

Due to the small number of participants in each condition, it was infeasible to do any sort of analysis of aptitude treatment interactions. With an increase in the number of participants and an increase in the number of problems solved during training, it would also be possible to look for an expertise reversal effect where examples are more effective for novices and less effective for experts [15].

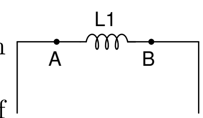
An interesting alternative would be a more integrated approach where the bottom-out hint of a hint sequence are replaced with the relevant worked-out example instead of replacing the entire hint sequence with an example. This is interesting from the standpoint of participant behavior, particularly "help abusers" [8]. It would be interesting to see if giving examples would reduce the help abuse behavior or encourage the participants to click through to see the example.

It could also be interesting to add worked-out examples to a more complicated Intelligent Tutoring Systems that can assess the utility of different actions and include worked-out examples as a possible action [16]. Worked-out examples could also be used in a system that could detect when dialog communication between the student and the system has broken down. When the communication is not achieving the systems goals, a worked-out example could be introduced to give the system and the student a common, grounded artifact to discuss.

A Appendix

A.1 Training Problem: IND1A

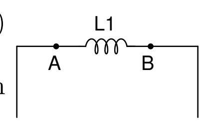
The current through an inductor L1 that has an inductance of 3.2 H is increased uniformly in time from 2.0 A in a direction from left to right to 3.5 A in a time of 3 s. Determine the emf (voltage) across the ends of the inductor (Vb-Va) throughout this time period.



A.2 Example: IND1

Problem Statement

Inductor L1 has a self-inductance of 10 H. The emf (voltage) between points A and B is held constant at 12 V. How much time does it take for the current through L1 to change from 2.4 A to -3.6 A?



Solution

Variables

T0 = current at 2.4 A to the right

T1 = current at 3.6 A to the left

t = duration of time from T0 to T1

IL0 = Current through L1 at time T0

IL1 = Current through L1 at time T1

VL = Voltage across L1 at time T0 to T1

L1 = inductance of L1

dIdt = rate of change of current through L1 at time T0 to T1

Equations

	Equation	Source
1.	L1 = 10 H	Given (This information is from the problem statement.)
2.	IL0 = -2.4 A	Given (This information is from the problem statement.)
3.	IL1 = 3.6 A	Given (This information is from the problem statement.)
4.	VL = 12 V	Given (This information is from the problem statement.)
5.	VL = -L1 * dIdt	Inductor EMF (VL = -L * dI/dt)
6.	dIdt = (IL1-IL0)/t	Constant (avg) Rate of Change (dI/dt = (I2-I1)/(t2-t1))
7.	VL = -L1 * (IL1-IL0)/t	6,5
8.	12 V = -10 H * (-2.4 A - 3.6 A) / t	7,4,1,2,3
9.	t = 5.0 s	8

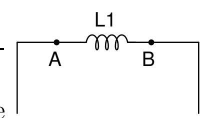
A.3 Post-Test Example

Instructions

In the following evaluation, you will be presented with a series of problem statements. You do not have to solve the problems! Your task will be to read the first problem statement and then decided which of the following two statements would be solved most similarly to the first one.

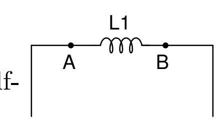
Problem Matching Task Question 1

The current through an inductor L1 that has an inductance of 3.2 H is increased uniformly in time from 2.0 A in a direction from left to right to 3.5 A in a time of 3 s. Determine the emf(voltage) across the ends of the inductor (Vb-Va) throughout this time period.

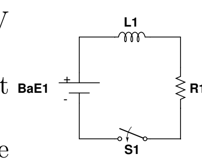


Which of the following two problems would be solved most similarly to the one above?

◯ Calculate the energy stored in a solenoid that has a self-inductance of 0.3 H when the current through it is 8.57 A.



◯ In the circuit shown, R1 = 2.0 ohm, L1 = 6.0 H and Vb = 12.0 V The switch is closed at T0. At the instant T1 when the current through the resistor reaches 4 A, what is the instantaneous rate of change of the current through the inductor L1? What is the energy in the inductor at this time?



References

1. Bloom, B.S.: The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher **13** (1984) 4–16
2. Sweller, J., Cooper, G.A.: The use of worked examples as a substitute for problem solving in learning algebra. Cognition and Instruction **2**(1) (1985) 59–89
3. Cooper, G., Sweller, J.: Effects of schema acquisition and rule automation on mathematical problem-solving transfer. Journal of Educational Psychology **79**(4) (1987) 347–362
4. Brown, D.E.: Using examples and analogies to remediate misconceptions in physics: Factors influencing conceptual change. Journal of Research in Science Teaching **29**(1) (1992) 17–34
5. Catrambone, R.: Aiding subgoal learning - effects on transfer. Journal of Educational Psychology **87**(1) (1995) 5–17
6. Koehler, M.J.: Designing case-based hypermedia for developing understanding children's mathematical reasoning. Cognition and Instruction **20**(2) (2002) 151–195
7. VanLehn, K., Lynch, C., Schulze, K., Shapiro, J.A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., Wintersgill, M.: Andes physics tutoring system: Five years of evaluations. In Looi, G.M.C.K., ed.: Proceedings of the 12th International Conference on Artificial Intelligence in Education, Amsterdam, IOS Press (2005)
8. Aleven, V., Koedinger, K.R.: Investigations into help seeking and learning with a cognitive tutor. Papers of the AIED-2001 Workshop on Help Provision and Help Seeking in Interactive Learning Environments (2001) 47–58
9. LeFevre, J.A., Dixon, P.: Do written instructions need examples? Cognition and Instruction **3**(1) (1986) 1–30

10. Trafton, J.G., Reiser, B.J.: The contribution of studying examples and solving problems to skill acquisition. In: Proceedings of the 15th Annual Conference of the Cognitive Science Society, Hillsdale: Lawrence Erlbaum Associates, Inc. (1993) 1017–1022

11. Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., Glaser, R.: Self-explanations: How students study and use examples in learning to solve problems. Cognitive Science **13**(2) (1989) 145–182

12. Chi, M.T.H., VanLehn, K.A.: The content of physics self-explanations. Journal of the Learning Sciences **1**(1) (1991) 69–105

13. VanLehn, K., Johns, R.M.: Better learners use analogical problem solving sparingly. In Utgoff, P.E., ed.: Machine Learning: Proceedings of the Tenth Annual Conference, San Mateo, CA, Morgan Kaufmann Publishers (1993) 338–345

14. Dufresne, R.J., Gerace, W.J., Hardiman, P.T., Mestre, J.P.: Constraining novices to perform expertlike problem analyses: Effects on schema acquisition. Journal of the Learning Sciences **2**(3) (1992) 307–331

15. Kalyuga, S., Ayres, P., Chandler, P., Sweller, J.: The expertise reversal effect. Educational Psychologist **38**(1) (2003) 23–31

16. Murray, R.C., VanLehn, K.: Dt tutor: A decision-theoretic, dynamic approach for optimal selection of tutorial actions. In Gauthier, Frasson, VanLehn, eds.: Intelligent Tutoring Systems: 5th International Conference. Volume 1839 of Lecture Notes in Computer Science., Montreal, Canada, Berlin: Springer (2000) 153–162

Hint Sequence 1:

```
T: You should finish entering all of the useful given quantities in
   the problem.  Why don't you work on entering the given value of the
   inductance of L1.
        Explain further OK


T: You can find the value of the inductance of L1 in the problem
   statement.


T: The value of the inductance of L1 is given as 3.2 H.


T: Enter the equation L1 = 3.2 H.
        OK
```

Hint Sequence 2:

```
T: Now that you have stated all of the given information, you should
   start on the major principles. What quantity is the problem seeking?
S: The Voltage across L1 at time T0 to T1.


T: That's right. What is the first principle application that you
   would like to work on? Hint: this principle application will
   usually be one that mentions the sought quantity
   explicity. Therefore it's equation may contain the sought quantity
   that the problem seeks.
S: V= -L*dI/dt   Inductor EMF
```

Fig. 2. These are two example of hint sequences that Andes provides students during problem solving. Andes lines are designated with a "T:" for tutor and the student responses are designated with a "S:." The first one shows the student requesting every hint in the sequence ending with the bottom out hint. The second one shows that the hints provide principle information. Students enter responses by selecting links in the dialog window or through a menu interface.

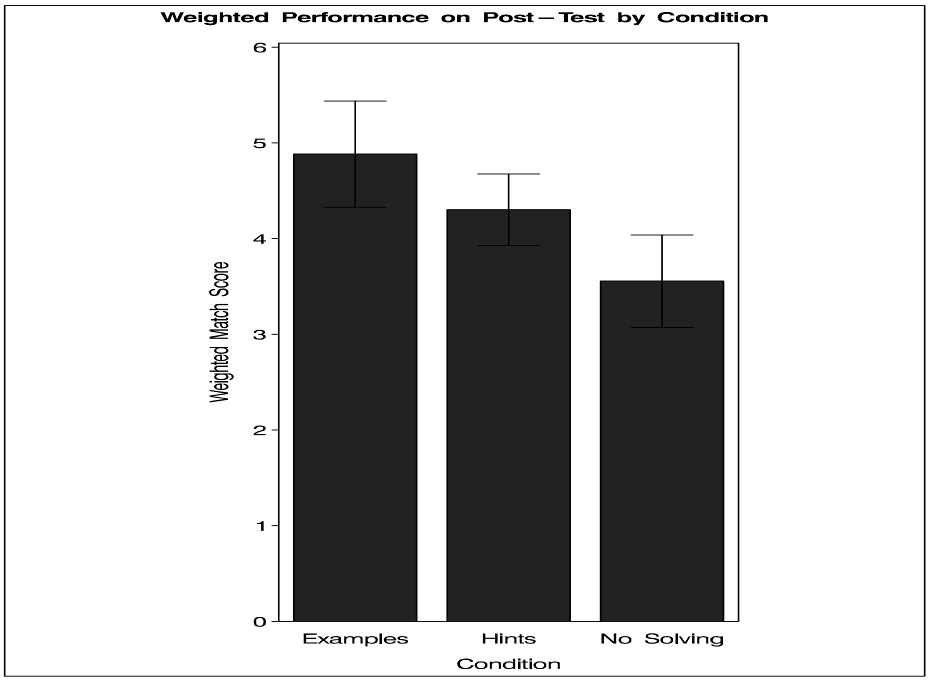


Fig. 3. Mean weighted post-test score by condition with whiskers showing the standard error of the mean.

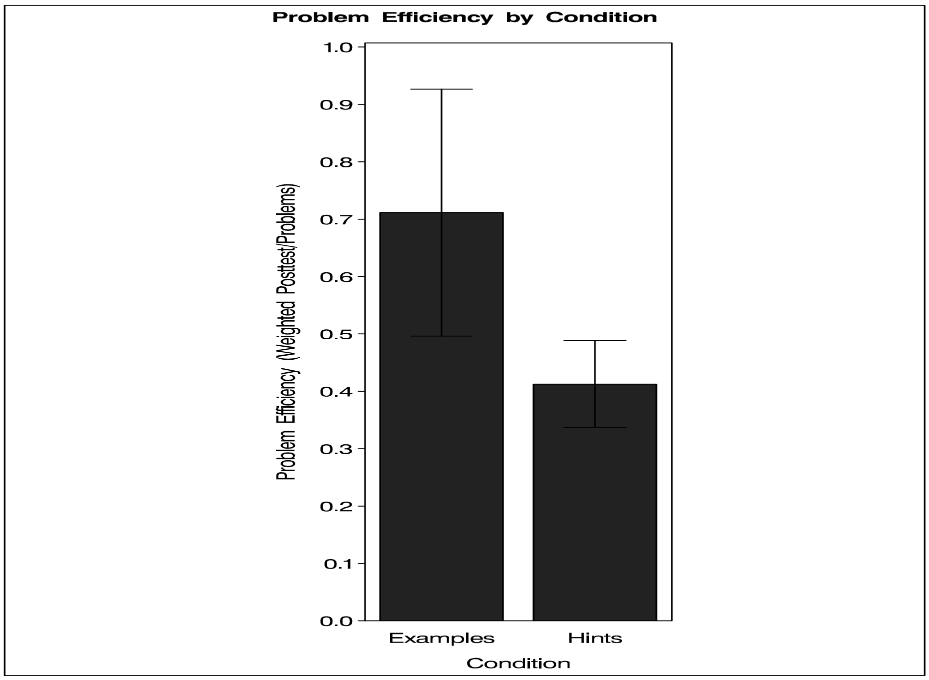


Fig. 4. Problem efficiency by condition where problem efficiency is the weighted post-test score divided by the number of problems solved during training with whiskers showing the standard error of the mean.

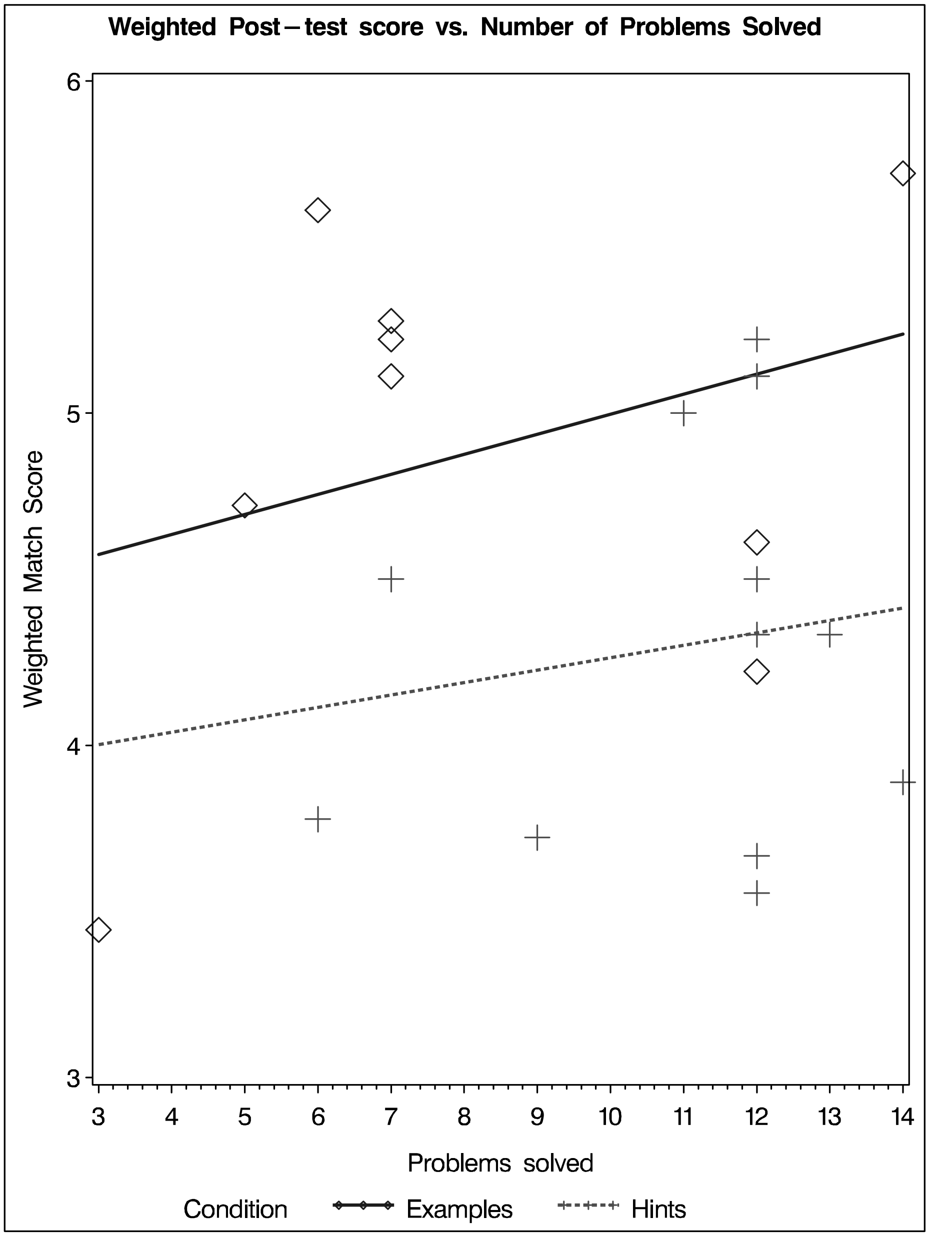


Fig. 5. Weighted post-test score versus number of problems solved with a fitted regression lines for the hints and examples conditions.