

# When Is It Best to Learn with All Worked Examples?

Bruce M. McLaren and Seiji Isotani

Carnegie Mellon University, Human Computer Interaction Institute,  
5000 Forbes Avenue, Pittsburgh, Pennsylvania  
{bmclaren, sisotani}@cs.cmu.edu

**Abstract.** Worked examples have repeatedly demonstrated learning benefits in a range of studies, particularly with low prior knowledge students and when the examples are presented in alternating fashion with problems to solve. Recently, worked examples alternating with intelligently-tutored problems have been shown to provide at least as much learning benefit to students as all tutored problems, with the advantage of taking significantly less learning time (i.e., more efficiency) than all tutored problems. Given prior findings, together with the prevailing belief that students should be prompted to actively solve problems after studying examples, rarely have *all* worked examples been tried as a learning intervention. To test the conventional wisdom, as well as to explore an understudied approach, a study was conducted with 145 high school students in the domain of chemistry to compare alternating worked examples / tutored problems, all tutored problems, and all worked examples. It was hypothesized that the alternating condition would lead to better results (i.e., better learning and/or learning efficiency) than either all examples or all tutored problems. However, the hypothesis was not confirmed: While all three conditions learned roughly the same amount, the all worked examples condition took significantly less time and was a more efficient learning treatment than either alternating examples/tutored problems or all tutored problems. This paper posits an explanation for why this (seemingly) surprising result was found.

**Keywords:** worked examples, intelligent tutors.

## 1 Introduction

The learning benefits of worked examples have been thoroughly researched and well documented [1]. A key theoretical reason often cited for the benefits of worked examples is cognitive load theory [2]. In particular, compared to problem solving, worked examples are believed to lessen *extraneous* load, which refers to the use of cognitive resources for mental processes, such as search. While search methods such as means-ends analysis are often critical to solving problems, such approaches exhaust the cognitive resources of students that could be used for learning. By providing learners with a worked-out solution to study, which worked examples do, the need for search is avoided and students can concentrate on building cognitive schemas, so they can more readily solve similar problems in the future.

Many studies have demonstrated the learning advantages of alternating worked examples with problems to solve (e.g., [3, 4, 5]). The learning benefits observed in

these studies appear to leverage a two-step learning process in line with cognitive load theory. First, it is helpful for a student, particularly one with low prior knowledge in the domain of interest, to review an example to lessen cognitive load and maximize initial learning. The cognitive schema created by the student while studying the example can then be used to, second, tackle an isomorphic problem to solve, i.e., one with similar structure and/or elements to the example. Instead of grappling with many new and unfamiliar details in solving the new problem, as well as searching through memory, the student can easily recall the similar, just-reviewed example while, at the same time, engage in active cognitive processing to (hopefully) strengthen their understanding of this type of problem and thus achieve deep learning [5].

More recent studies have investigated the benefits of alternating worked examples with intelligently tutored problems [6]. These empirical investigations differ from more traditional worked examples research by the inclusion of *tutored* problems to solve, which provide step-by-step guidance in the form of hints and error feedback and thus offer more scaffolding than ordinary problems. Tutored problems are a middle ground between worked examples and problem solving: they allow students, if they wish, to create worked examples from the problems (by e.g. drilling down to bottom-out hints) but, at the same time, students can actively attempt to solve problems. These recent studies also differ from earlier research in that they have mostly been conducted in the classroom, a decidedly more difficult environment to test learning interventions than the laboratory setting of most prior studies.

All of these recent studies have tested the hypothesis that replacing some tutored problems with worked examples will enhance student learning by reducing instructional time and/or increasing student learning, in terms of retention and transfer. For instance, Schwonke and colleagues [7], in two studies in the domain of geometry, found that a fading worked examples condition, one in which some tutored problems were replaced with examples that were, in turn, gradually replaced first by partially completed examples and, later, by fully-tutored problems, led to as much learning and transfer as a control condition of all tutored problems, yet in significantly less time. McLaren et al's findings [8] in three studies in the domain of chemistry corroborated the efficiency findings of Schwonke et al – more specifically, an alternating example-tutored problem condition yielded the same learning as a tutored-problems-only control, but with significantly better learning efficiency. In summary, it appears that adding worked examples to tutored problem solving helps learning, but the benefits are mostly in improving learning efficiency. Learning outcomes are generally not as significant as those reported in untutored problem solving research (e.g., [3, 4]), which may be explained, at least in part, by the tougher control condition all tutored problems presents.

Given these past findings, both in untutored and tutored problem solving research, together with the generally accepted two-step learning process discussed above and the cognitive load theory underlying it, it is not surprising that researchers have infrequently tested the benefits of presenting students with all worked examples. Why would we expect superior learning benefits with all worked examples when the active problem solving step – the second of the two steps, the step that reinforces the example and (possibly) leads to deep learning – is taken out of the equation? Because we were interested in exploring the never-tested all-worked examples condition (at least never tested within the tutored problem solving line of research) and (re-)testing

the generally-accepted two-step learning process of examples followed by problems to solve, we ran a study of the all-examples condition in the context of tutored problem solving. We were skeptical that all-worked examples could produce better learning outcomes or efficiency and formulated the following hypothesis.

*Alternating worked examples and tutored problem solving will lead to better learning (i.e., better learning retention and/or learning efficiency) than either all tutored problems or all worked examples.*

Given our own prior results [8], which showed better learning efficiency in alternating examples/tutored problems versus all tutored problem solving, as well as the preponderance of evidence supporting the advantages of the two-step learning process [3, 4, 5, 6], the first part of this hypothesis (i.e., *alternating examples/tutored problems > all tutored problems*) was already well supported. Despite some (but limited) evidence that all examples can be more effective for learning and more efficient in mental effort, at least as compared to all untutored problem solving [9, 10], our theory was that all examples might be faster than examples/tutored problems but likely at the expense of careful study and robust learning, thus hurting both learning outcome and efficiency, suggesting the second part of our hypothesis (i.e., *alternating examples/tutored problems > all worked examples*). On the other hand, an “in press” study, one that occurred more-or-less concurrently to ours (yet after our hypothesis was formulated), casts doubt on the notion that an alternating condition is better than all examples, at least with respect to regular, non-tutored problem solving. In this study all worked examples were *as good as* alternating examples/problem solving, with both conditions better than all problems, in terms of both lower cognitive load during learning and higher learning outcomes [11]. Thus, it is clear that the outcome of our study, which will now be described, was not obvious.

## 2 Method

**Participants.** One hundred and forty-five (145) high school students (67 female and 78 male) in four chemistry classes in three suburban high schools in three U.S. states (Pennsylvania, Massachusetts and New Jersey) participated. There were 11 additional students who scored 0 (or very nearly 0) on one or both of the two posttests and 3 more students who finished the delayed posttest much later than their classmates; all of these students were eliminated from consideration. The study materials were used as a replacement for normal lectures and class work on the topic of stoichiometry within the four high school classes, and the three participating teachers used the immediate and delayed posttests as class grades for their students.

**Materials and Procedure.** We conducted a between-subjects study with students randomly assigned to one of the three conditions shown in Table 1. Students in condition 1 (*exs/tps*) were presented with 5 alternating pairs of isomorphic examples and tutored problems, students in condition 2 (*all-tps*) were presented solely with 10 tutored problems corresponding to the same problems received in condition 1, and students in condition 3 (*all-exs*) were presented solely with 10 worked examples corresponding to the same problems received in condition 1. While the intervention materials varied by condition, all students were presented with the same consent form, pre-questionnaire, preparation videos, post-questionnaire, immediate posttest, and delayed posttest. The *n* of each condition is shown at the top of Table 1.

**Table 1.** Study Design with Three Conditions.

<b>Condition 1:</b> <i>exs/tps (n=45)</i>	<b>Condition 2:</b> <i>all-tps (n=51)</i>	<b>Condition 3:</b> <i>all-exs (n=49)</i>
Consent Form		
Pre-Questionnaire		
Five Preparation Videos		
<u>Intervention Materials:</u> 10 problems in 5 isomorphic pairs: <i>1st in each pair – example</i> <i>2nd in each pair – tutored problem</i> (3 content videos interspersed)	<u>Intervention Materials:</u> 10 problems in 5 isomorphic pairs, <i>all tutored problems</i> (3 content videos interspersed)	<u>Intervention Materials:</u> 10 problems in 5 isomorphic pairs, <i>all worked examples</i> (3 content videos interspersed)
Post-Questionnaire		
Immediate Posttest – 8 problems (4 near transfer; 4 conceptual)		
Delayed Posttest (one week later) – 8 problems (4 near transfer; 4 conceptual)		

All materials were completed online, within a web browser, in the top-down order shown in Table 1. Students used school-provided computers and headphones, so they could privately listen to the videos. All participants were given user-IDs and passwords that allowed them to logoff and log back on whenever desired, including outside of the classroom.

Because of the usual difficulties in tightly controlling classroom time, the study materials of Table 1 were tackled mostly, but not exclusively, during teacher-monitored classroom time. In a few cases, due to absences or insufficient classroom time, the consent form, questionnaires, videos, and intervention materials were completed outside of classroom time, either at school or home. However, all of the posttests were taken in class. The immediate posttest was administered in the class following completion of the intervention materials and the delayed posttest was administered one week later. Each posttest took 45 to 60 minutes to complete and all of the materials in Table 1 took students between 150 and 240 minutes to complete.

The pre-questionnaire contained basic demographic questions (e.g., gender), as well as self-assessment questions about the student's prior knowledge of chemistry (e.g., "I know what the 2 stands for in H<sub>2</sub>O", "I know what Na stands for," "Rate your overall knowledge of chemistry, from 1 ('Far below average') to 5 ('Highly above average')"). The pre-knowledge questions were scored between 1 and 15 and all students who scored below the calculated mean of 9.95 were classified as "low prior knowledge learners," while all others were classified as "high prior knowledge learners." Note that we did not administer a pretest, in favor of the self-assessment questions, to avoid testing effects, the phenomenon in which a test can help students learn after they have already pre-studied material [12]. As pupils in a chemistry class, our population of subjects may have been exposed to similar or related materials.

After completing the pre-questionnaire, the students were presented with five videos to prepare them for working with the materials, including a review of how to use the online materials, a review of significant figures, and an overview of stoichiometry problem solving. Next, the students worked on the intervention materials that were specific to their condition, as shown in Table 1. Short (1 to 4 minute) videos were interspersed throughout the materials, presenting various background materials on chemistry concepts relevant to stoichiometry (e.g., molecular weight, dimensional analysis).

Fig. 1 shows the Stoichiometry Tutor, developed using the Cognitive Tutor Authoring Tools [13]. To solve the stoichiometry problems, students must understand basic chemistry concepts, such as molecular weight, and be able to solve algebraic chemistry equations. The student can request hints by selecting the “Hint” button in the upper right-hand corner of the interface. The hints the tutor gives provides progressively more information for solving the problem, with the last hint on each step providing the final answer for that step (a “bottom-out hint”, Fig. 1 is an example of such a hint). If the number typed (or unit or substance or reason selected) is correct, the typed (or selected) information appears in a green font. If it is incorrect, it appears in red. The tutor also provides context-specific error messages when the student makes a mistake during problem solving.

The screenshot shows the Stoichiometry Tutor interface. At the top right is a "Hint" button. Below it is the "Problem Statement" which asks if a water sample with 0.58 moles of arsenite (AsO<sub>2</sub><sup>-</sup>) in 100 kiloliters (100 kL) is safe, given a WHO limit of 0.000014 grams of arsenite (AsO<sub>2</sub><sup>-</sup>)/L solution. The molecular weight of arsenite (AsO<sub>2</sub><sup>-</sup>) is 106.9 g AsO<sub>2</sub><sup>-</sup> / mol AsO<sub>2</sub><sup>-</sup>.

The "Problem" section contains a calculation grid with columns for "#", "Units", and "Substance". The first row shows 0.58 mol of AsO<sub>2</sub><sup>-</sup>. The second row shows 100 kL of solution. The third row shows a conversion factor of 1 kL solution / 1000 L. The result column shows 0.58 g of AsO<sub>2</sub><sup>-</sup>.

Below the grid are four "Reason" dropdown menus, with the first one set to "Given Value". A "Done" button is to the right. At the bottom, a green "Hint" box states: "Hint: Do you recall that 1000 L is equivalent to 1 kL? Thus, you may want to type 1000 here." A "Next" button and a "get previous hint" button are also visible.

Fig. 1. The Stoichiometry Tutor and an example of a bottom-out hint

The worked examples used in Conditions 1 and 3 are implemented as videos of an expert solving and narrating problems with the Stoichiometry Tutor shown in Fig. 1. To prompt reflection and review of the example, the student is asked, after watching a worked example video, to answer several (3 to 5) self-explanation (SE) questions, through multiple-choice, pull-down menus. Examples of SE questions for the worked example version of the problem in Fig. 1 are: “Our goal in this problem was to convert from moles AsO<sub>2</sub><sup>-</sup> per kiloliters solution to grams AsO<sub>2</sub><sup>-</sup>: (a) true, (b) false.”, “We used the unit conversion term 1 kL solution / 1000 L to convert: (a) the numerator of the given value from moles to grams, (b) the denominator of the given value from kiloliters to liters, (c) the numerator of the given value from moles to liters, (d) the denominator of the given value from liters to kiloliters.” If the student incorrectly answers an SE question, the answer turns red. The student cannot proceed to the next problem until they answer all SE questions correctly.

After completing all of the intervention videos and problems, the participants were prompted to respond to a web-based questionnaire that asked about the helpfulness and the usability of the tutor. Finally, the students took an immediate posttest and, one week later, a delayed posttest. The two posttests were isomorphic to one another (A and B), each containing eight problems, 4 of which were of the same type and had the same user interface as the intervention problems (like Fig. 1 but without hints or error

feedback; near transfer problems) and 4 of which were more conceptual questions. The order of the A and B tests was counterbalanced (i.e., ½ of the participants received A as the immediate test and B as the delayed test and vice versa).

Test scores were calculated by assigning a score per problem (i.e., dividing the number of correct steps the student took on a single problem by the total number of possibly correct steps for that problem), adding the scores of all eight problems together, and dividing by 8.

### 3 Results and Discussion

The results comparing the three conditions according to the dependent variables (DVs) immediate and delayed posttest performance are shown in Table 2. ANCOVAs on the DVs immediate and delayed posttest performance, with prior knowledge as a covariate, were run, with post-hoc comparisons between each pair of conditions, including a calculation of effect size (Cohen's *d*). As can be seen, there were no significant differences in learning between the conditions on either immediate or delayed posttest performance. Separate analyses of the low and high prior knowledge students (not shown in Table 2), as defined by the 9.95 threshold discussed earlier, also did not exhibit significant effects. However, for the low prior knowledge group the ex/tps condition compared to both the all-tps and all-exs conditions reached medium, but not significant, effect sizes for both the immediate and delayed posttests. This is at least a hint toward past results that have shown, for low prior knowledge students, alternating worked examples with problems can be advantageous (e.g., [4]).

**Table 2.** Comparison according to the DVs imm. posttest and delayed posttest performance

Dependent Variable	C1: exs/tps mean (sd)	C2: all-tps mean (sd)	C3: all-exs mean (sd)	<i>p</i>	Cohen's <i>d</i> C1 v. C2	Cohen's <i>d</i> C1 v. C3	Cohen's <i>d</i> C2 v. C3
Imm. Posttest	0.54 (0.18)	0.53 (0.16)	0.54 (0.19)	0.84	0.07	-0.01	-0.08
Del. Posttest	0.62 (0.19)	0.61 (0.16)	0.58 (0.23)	0.27	0.05	0.15	0.11

\* - significant difference,  $p < 0.05$

The results comparing the three conditions according to the DVs' learning time, learning efficiency on the immediate posttest, and learning efficiency on the delayed posttest, are shown in Table 3<sup>1</sup>. Here, the learning time of the all-exs condition was found to be significantly shorter than either the exs/tps or all-tps conditions, with large effect sizes in both cases. The learning efficiency of the immediate posttest of the all-exs condition was also significantly higher than either of the other two conditions, with medium to high effect size. The learning efficiency of the delayed posttest of the all-exs condition was marginally significantly higher than the alternating condition and significantly higher than the all-tps condition, with medium effect sizes. Notice that the comparison of the alternating condition to the all-tps condition, unlike our prior studies with the same tutors and population [8], was not significant on any of the DVs.

<sup>1</sup> Learning efficiency was calculated per subject as  $z\text{-score}(\text{test score}) - z\text{-score}(\text{instructional time})$  with  $z\text{-score} = (\text{value} - \text{mean}) / \text{sd}$ . This measure is a simplified (but mathematically sound) version of the quantitative model of efficiency first described in [14].

**Table 3.** Comparison according to the DVs learning time, learning efficiency (imm. posttest), and learning efficiency (del. posttest).

Dependent Variable	C1: exs/tps mean (sd)	C2: all-tps mean (sd)	C3: all-exs mean (sd)	<i>p</i>	Cohen's <i>d</i> C1 v. C2	Cohen's <i>d</i> C1 v. C3	Cohen's <i>d</i> C2 v. C3
Learning Time	59.4 (18.0)	63.9 (20.8)	43.0 (12.6)	< 0.0001	-0.23	1.06 *	1.21 *
Learning Eff. (Imm. Posttest)	-0.18 (1.44)	-0.48 (1.54)	0.66 (1.11)	< 0.0001	0.20	-0.66 *	-0.85 *
Learning Eff. (Del. Posttest)	-0.13 (1.44)	-0.41 (1.57)	0.54 (1.17)	0.003	0.18	-0.51 #	-0.68 *

\* - significant difference,  $p < 0.05$  # - marginally significant difference,  $0.05 < p < 0.1$

Our hypothesis was therefore not confirmed: The exs/tps condition did not lead to better learning, either in retention or efficiency, than either the all-exs or all-tps conditions. In fact, the all examples condition showed learning efficiency benefits compared to the other two in this study. Why might this have occurred? The answer may lie in the two-step learning process discussed earlier. The worked examples of this study may provide students with an opportunity to take both of the steps within the context of one example. Students first study the video example and are then encouraged to actively process the material through the follow-up SE questions, which cannot be skipped and must be answered correctly to proceed. While self-explanation is not problem solving, it is implemented in this case in a manner that may trigger a sort of “mental” problem solving. Furthermore, the isomorphic second worked example of each pair may strongly reinforce learning of particular problem types, through a second cycle of the two-step process. In short, while the all-exs condition did not lead to a better learning outcome than the other conditions, it may have promoted a learning model similar to what has worked for alternating examples and tutored problem solving, leading to as much learning yet in a much faster way.

Perhaps even more surprising is the fact that, when comparing exs/tps to all-tps, we did not replicate the significant learning time benefit achieved in 3 of 3 prior studies (effect sizes 1.02, 0.59, 0.54 vs. 0.23 in this study, i.e., viewed in positive direction, instead of the -0.23 in Table 3) or the learning efficiency benefit achieved in 2 of 3 prior studies (effect sizes 0.75, 0.39, 0.56 vs. 0.20 on immediate posttest in this study) [8], even though the same materials and general population were used. We currently have no explanation for this outcome and will continue to explore the data for clues.

## 4 Conclusion

Our study produced surprising results. While our hypothesis was not confirmed, including replication of our own past work, we made an interesting discovery: at least sometimes, under some conditions, students can benefit the most by learning strictly with worked examples, at least with respect to conserving their time. The worked examples of our study, however, were not static, conventional examples; rather, they were “modeling” examples [15] – live, narrated videos – that were followed by prompted self-explanation questions that had to be answered correctly by the student in order to move on. At least in the context of our domain and our materials, such a type of worked example, which trades off between example study and active problem processing, led to the best results, if not with respect to learning gains than at least with respect to learning efficiency. In other words, our study showed that learning

from what might be called “interactive” worked examples may sometimes be a better choice than static worked examples, tutored problems, or problems to solve.

**Acknowledgments.** The Pittsburgh Science of Learning Center (PSLC), NSF Grant # 0354420, provided support for this research. Dave Yaron and Mike Karabinos assisted in developing the stoichiometry materials used in the study. Tamara van Gog reviewed an early version of this paper and provided valuable comments.

## References

1. Atkinson, R.K., Derry, S.J., Renkl, A., Wortham, D.: Learning From Examples: Instructional Principles from the Worked Examples Research. *Review of Educational Research* 70, 181–214 (2000)
2. Sweller, J., Van Merriënboer, J.J.G., Paas, F.: Cognitive Architecture and Instructional design. *Educational Psychology Review* 10, 251–295 (1998)
3. Mwangi, W., Sweller, J.: Learning to Solve Compare Word Problems: The Effect of Example Format and Generating Self-Explanations. *Cog. and Inst.* 16, 173–199 (1998)
4. Kalyuga, S., Chandler, P., Tuovinen, J., Sweller, J.: When problem solving is superior to studying worked examples. *Journal of Educational Psychology* 93, 579–588 (2001)
5. Sweller, J., Cooper, G.A.: The use of Worked Examples as a Substitute for Problem Solving in Learning Algebra. *Cog. and Inst.* 2, 59–89 (1985)
6. Salden, R.J.C.M., Koedinger, K.R., Renkl, A., Alevén, V., McLaren, B.M.: Accounting for Beneficial Effects of Worked Examples in Tutored Problem Solving. *Educational Psychology Review* 22(4), 379–392 (2010)
7. Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Alevén, V., Salden, R.J.C.M.: The Worked-Example Effect: Not an Artefact of Lousy Control Conditions. *Computers in Human Behavior* 25, 258–266 (2009)
8. McLaren, B.M., Lim, S., Koedinger, K.R.: When and How Often Should Worked Examples be Given to Students? New Results and a Summary of the Current State of Research. In: *Proc. of the 30th Annual Conf. of the Cog. Sci. Soc.*, pp. 2176–2181 (2008)
9. Van Gerven, P.W.M., Paas, F., Van Merriënboer, J.J.G., Schmidt, H.G.: Cognitive Load Theory and Aging: Effects of Worked Examples on Training Efficiency. *Learning and Instruction* 16, 154–164 (2002)
10. Van Gog, T., Paas, F., Van Merriënboer, J.J.G.: Effects of Process-Oriented Worked Examples on Troubleshooting Transfer Performance. *Learn. and Inst.* 18, 211–222 (2006)
11. Van Gog, T., Kester, L., Paas, F.: Effects of Worked examples, Example-Problem, and Problem-Example Pairs on Novices’ Learning. *Contemporary Ed. Psychology* (in press)
12. Roediger III, H.L., Karpicke, J.D.: The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psych. Science* 1(3), 181–255 (2006)
13. Alevén, V., McLaren, B.M., Sewall, J., Koedinger, K.R.: A New Paradigm for Intelligent Tutoring Systems: Example-Tracing Tutors. *Int’l J. of AIED* 19(2), 105–154 (2009)
14. Paas, F., Van Merriënboer, J.J.G.: The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. *Human Factors* 35(4), 737–743 (1992)
15. Van Gog, T., Rummel, N.: Example-Based Learning: Integrating Cognitive and Social-Cognitive Research Perspectives. *Educational Psychology Review* 22, 155–174 (2010)