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.

When and How Often Should Worked Examples be Given to Students? New Results and a Summary of the Current State of Research

Bruce M. McLaren, Sung-Joo Lim, and Kenneth R. Koedinger
(bmclaren@cs.cmu.edu, sungjol@andrew.cmu.edu, koedinger@cs.cmu.edu)
Human-Computer Interaction Institute,
5000 Forbes Avenue,
Carnegie Mellon University,
Pittsburgh, PA, 15213-3891 United States

Abstract

Our work explores the *assistance dilemma*: when should instruction provide or withhold assistance? In three separate but very similar studies, we have investigated whether worked examples, a high-assistance approach, studied in conjunction with tutored problems to be solved, a mid-level assistance approach, can lead to better learning. Contrary to prior results with *untutored* problem solving, a low-assistance approach, we found that worked examples alternating with isomorphic tutored problems did not produce more learning gains than tutored problems alone. On the other hand, the examples group across the three studies learned more efficiently than the tutored-alone group; the students spent 21% less time learning the same amount of material. Practically, if these results were to scale across a 20-week course, students could save 4 weeks of time – yet learn just as much. Scientifically, we provide an analysis of a key dimension of assistance: when and how often should problem solutions be given to students versus elicited from them? Our studies, in conjunction with past studies, suggest that on this example-problem dimension mid-level assistance may lead to better learning than either lower or higher level assistance. While representing a step toward resolving the assistance dilemma for this dimension, more studies are required to confirm that mid-level assistance is best and further analysis is needed to develop predictive theory for what combinations of assistance yield the most effective and efficient learning.

Keywords: Instruction and Teaching, Learning, Skill acquisition and learning

Introduction

Building on past notions like "zone of proximal development" (Vygotsky, 1978) and cognitive apprenticeship (Collins, Brown, & Newman, 1990), the *assistance dilemma* (Koedinger & Aleven, 2007) characterizes a long-standing unsolved problem in the learning sciences: when should instruction provide students with assistance and when should it be withheld? Some researchers have argued for providing maximal assistance (e.g., Kirschner, Sweller, & Clark, 2006) while others argue for minimal assistance (e.g., Steffe & Gale, 1995).

In three studies in the domain of chemistry, we have explored the assistance dilemma, investigating whether two instructional devices – worked examples and personal/polite language – can provide learning support beyond what is provided by an intelligent tutoring system (McLaren *et al*, 2006; 2007). In this paper we focus exclusively on the worked examples aspect of our studies. More specifically, we summarize the McLaren *et al* results in experimenting with an intelligent tutor supplemented with worked examples (a combination that has only recently been investigated) and discuss new analyses of these three studies. The worked example principle, as stated in Clark & Mayer (2003) is: "Replace some practice problems with worked examples", i.e., provide students with an alternating combination of worked examples and problems. The theory behind the principle is that human working memory, which has a limited capacity, is taxed by strictly solving problems, which requires thinking, such as the setting of subgoals. Such mental work consumes cognitive resources that could be better used for learning (Sweller, Van Merriënboer, & Paas, 1998). The rationale, then, is that worked examples free those resources for learning processes, in particular, the induction of (or modifications to) knowledge components.

But then why *mix* worked examples and problem solving, as suggested by the worked example principle? The theory seems to suggest that worked examples provided alone, a high-assistance approach, would be best for learning. What does empirical research say about this theory and the combination of worked examples and problem solving?

One way of answering this question is to evaluate past, representative studies along an *example-problem dimension of assistance*, which represents different levels of assistance that students may receive while learning (see Figure 1). Arguably, problem solving with no tutoring is the least assistance approach (level "1" in Figure 1), followed by problem solving with tutoring ("2"), worked examples with no explanation of individual problem-solving steps ("3"), and, finally, the highest assistance case is worked examples with explanations of individual steps ("4"). The vertical arrows next to each of the studies on the dimension of assistance show the conditions compared in that study. Thick arrows indicate precise conditions on the continuum (e.g., the Paas, 1992 study had one condition which was precisely level 1) or contiguous, combination conditions (e.g., the Schwonke *et al*, 2007 study had one condition which alternated assistance between levels 2 and 3), while thin arrows denote noncontiguous, combination conditions (e.g., the Paas, 1992 study had a second condition which alternated levels 1 and 3).

Lovett's study (bottom of Figure 1) compared all four levels of assistance[1] and found that problem solving without tutoring was best, with superior near and far transfer gains (indicated by the "+" and "#" signs), while worked examples with explanations, on the other end of the spectrum, also led to superior far transfer gains (indicated by the "+"), as compared to the middle two conditions (Lovett, 1992).

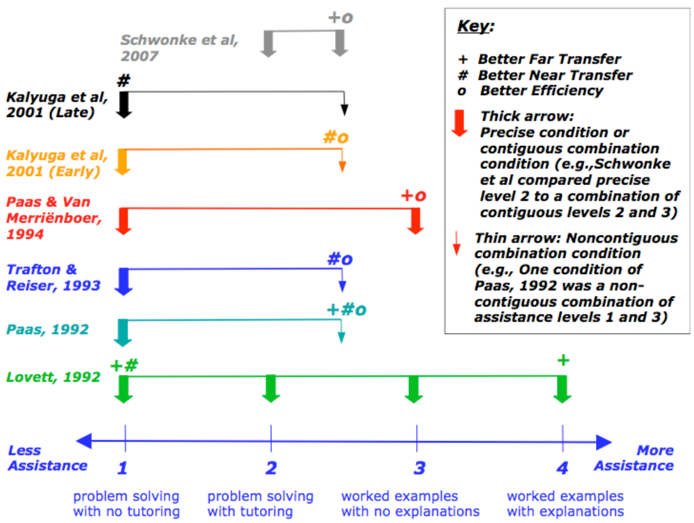

Figure 1: The example-problem dimension of assistance and a variety of studies that have compared different levels of assistance, e.g., Paas & Van Merriënboer, 1994 compared problem solving with no tutoring to worked examples with no explanations, finding better near and far transfer for the latter.

Paas found that students who studied eight unexplained worked examples and solved four untutored problems (a mixed condition indicated by the thin arrow pointing between "2" and "3") worked for less time and scored higher on both near and far transfer tests than students who solved all 12 problems (Paas, 1992). Trafton and Reiser (1993) compared problem solving with no tutoring to interleaved worked examples and problem solving with no tutoring. They found statistically significant near transfer learning gains and learning efficiency for the alternating condition. Paas and Van Merriënboer (1994) compared problem solving with no tutoring to all worked examples with no explanations, finding the all examples condition to be significantly better in both far transfer learning and efficiency[2]. Kalyuga and colleagues (2001) compared untutored problem solving with alternating unexplained examples and untutored problem solving in an extended experiment with multiple stages and training sessions. They initially found a significant difference in normal learning gains and efficiency in favor of the mixed examples / problem solving condition (indicated as the "Early" study on the dimension of assistance) but, as students gained more expertise through training sessions, a significant near transfer (but not efficiency) advantage to problem solving was identified (indicated as the "Late" study).

More recently, researchers have compared the region of this dimension of assistance that represents *tutored problem solving* with other forms of assistance. For instance, the study that Schwonke and colleagues (2007) conducted compared tutored problem solving with alternating worked examples and tutored problem solving. They got a null effect for normal learning gains in two separate studies, but learning was more efficient in both studies with transfer learning found in the second study. The studies discussed in this paper are similar to the Schwonke *et al* work, in that they compare alternating worked examples and tutored problem solving with tutored problem solving alone, but differ in that Schwonke *et al* explicitly leveraged the results of Kalyuga *et al* (2001) by "fading" examples from the materials, as students gained expertise. No example fading was done in the studies reported in this paper.

Why Isn't the Science Done?

Taken together, the studies in Figure 1 give rise to a couple observations – and scientific questions – about the dimension of assistance and the assistance dilemma. First, notice that the results in Figure 1 are not definitive on the issue of whether more or less assistance is beneficial to learning. For instance, the Lovett study demonstrates that *both* a low and a high assistant approach could be beneficial, and the Kalyuga studies suggest that assistance should decline over time, as subjects gain expertise. Thus, there is clearly room for continued studies comparing levels of the example-problem dimension of assistance. Second, as already noted, until recently there had been little study of the comparative contributions of learning with *intelligent tutored* problem solving and other forms of assistance. Tutored problem solving is a mid-level assistance approach that provides more assistance than untutored problem solving but somewhat less than worked examples. Only the Schwonke *et al* study, as well as our own, have explored the combination of tutored problems and worked examples. Finally, and somewhat contrary to the first observation, notice that most of the results, beginning with Paas (1992), indicate a tendency for *mid-level* assistance being most beneficial to learning, and in particular the approach of alternating worked examples with problem solving. In fact, the worked examples principle is based on these findings (Clark & Mayer, 2003). Thus, it appears *the example-problem dimension of assistance may be represented as an inverted-U*, in which the mid-level approaches yield the greatest learning benefits, while the lesser and greater assistance approaches yield somewhat lesser benefits (at least for the average student). A hypothesis that arises from

[1] Note, however, that "problem solving with tutoring" was not *intelligent* tutoring, but rather elaborated explanations provided by a human experimenter during problem solving.

[2] It is worth noting that in this study – as well as others, such as (Lovett, 1992; Trafton & Reiser, 1993) – examples and/or solutions were provided in the problems-only condition *after* a student unsuccessfully completes a problem. Thus, there is an element of "worked examples" even in the pure problem solving condition.

these observations – and the one we are interested in and have tested in the studies reported in this paper – is:

> *The interleaving of worked examples with problems* ***supported by an intelligent tutor*** *will further improve learning beyond the benefits of the tutor itself*

Does the assistance provided by an intelligent tutor possibly replace the assistance of worked examples? Consider, for example, that a tutor could be seen as a way of converting a problem into an example by providing the next step, as a hint, when the student is stuck. In short, exploring the combined affect of worked examples and tutors – and how the two types of assistance differ from and/or complement one another – is still an open scientific question.

In addition to exploring the above hypothesis and continuing to flesh out the example-problem dimension, continued worked examples studies are scientifically important because the worked example principle relies primarily on short-duration lab studies; it has rarely been tested in real classrooms over longer durations[3]. That is, most past studies have lacked ecological validity, since subjects were paid, worked with content outside a real academic curriculum, and studied the materials for short periods of time, often for less than an hour. The studies discussed in this paper were done for class credit (except for the first study), covered topics that are part of an intro to chemistry course, and took students from 1.5 to 6.5 hours to complete all materials (i.e., pretest, tutors, worked examples, videos, posttest, and questionnaires).

The Stoichiometry Tutor and Examples

Our studies involved the learning of stoichiometry and the use of the Stoichiometry Tutor (McLaren *et al*, 2006). Solving a stoichiometry problem involves understanding basic chemistry concepts (e.g., the mole, unit conversions) and applying those concepts in solving equations of ratios. The student must fill in the terms of an equation, correctly cancel numerators and denominators, provide reasons for each term (e.g., "Molecular Weight"), and calculate and fill in a final result.

Applying the principles of cognitive tutoring (Anderson *et al*, 1995), the tutor provides the student with hints on request and also provides context-specific error messages when the student makes a mistake. For more description of both the stoichiometry problems and the Stoichiometry Tutor itself, see (McLaren *et al*, 2006).

Worked examples in the studies are Flash videos in which a narrator solves a stoichiometry problem using the Stoichiometry Tutor, describing each of the steps taken. (Note that worked examples are higher assistance than tutor use, as intermediate steps and answers are provided in the worked examples without the student asking for hints.) After watching the video the student is prompted with 3 to 5 multiple-choice, self-explanation questions. Their responses are "graded" (i.e., right or wrong) and the student cannot proceed until they have correctly answered all of the self-explanation questions. Self-explanation is a robust learning principle that has been shown in many studies to promote deeper learning, beginning with the work of Chi *et al* (1989).

Study Design and Procedure

For all three studies a 2x2 factorial design was employed. The independent variable of primary interest in this paper is *Worked Examples*, with one level being *Tutored Alone* and the other *Worked Examples + Tutored*. In the former condition, which will be referred to as the "Problems Condition" henceforth, subjects only solved problems with the tutor; no worked examples were presented, as shown in the left column of Table 1. In the latter condition, which will be referred to as the "Examples Condition" and which is illustrated in the right column of Table 1, subjects alternated between observation and prompted self-explanation of a worked example (as previously described) and solving of an isomorphic problem with the aid of the Stoichiometry Tutor (i.e., Study Problems #1 and #2 are isomorphic to one another, #3 and #4 are isomorphic, and so on). Isomorphic problem solutions have the same number, type, and order of terms. The second independent variable, *personalization*, with one level personal problem statements the other impersonal problem statements, has not and will not be further discussed, since it is not the focus of the current paper. Discussion of this variable and findings related to it can be found in McLaren et al (2006; 2007).

All instructional materials were provided via the Internet. All subjects were given pre- and post-questionnaires, requesting demographic information, chemistry background, and – in the post-questionnaire – assessment of the tutors. All subjects were also given online pre and posttests, with the problems on the posttest isomorphic to the pretest problems. All pre and posttest problems involved the same type of problems as the study problems. The subjects worked on the 10 study problems, presented according to the conditions of Table 1, with the Problems Condition working only on tutored problems and the Examples Condition working on alternating (and isomorphic) examples and tutored problems (ala Trafton and Reiser (1993)). Instructional videos on chemistry content were intermingled with the study problems in both conditions.

All individual steps taken by the students in the pretest and posttest were logged and automatically marked as correct or incorrect. A normalized score between 0 and 1.0 was calculated for each student's pre and posttest by dividing the number of correct steps by the total number of possibly correct steps. Pretest scores indicated that students were balanced across conditions (except for low pretest scores in the Problems Condition of study 2, see Figure 2).

Table 2 summarizes the N value, target populations, and noteworthy characteristics of the three studies.

[3] One exception is the Kalyuga *et al* study (2001). Although not a classroom study with intelligent tutors, they tested over periods of greater than 6 hours.

Table 1. Study Design for the independent var. Worked Examples[4]

Problems Condition (i.e., Tutored Alone)	Examples Condition (i.e.,Worked Examples + Tutored)*
Pre-Questionnaire	< Same as on left >
Videos: Introduction to Stoich Study, Intro to Pretest User Interface	< Same as on left >
5 Pretest Problems	< Same as on left >
Videos: Intro to Study problems, Stoichiometry Problem Solving Strategy, Dimensional Analysis & Avogadro's #, Significant Figures	< Same as on left >
Study Problem # 1	Worked Ex. of Problem #1
Study Problem # 2	< Same as on left >
Video: Molecular Weight	< Same as on left >
Study Problem # 3	Worked Ex. of Problem #3
Study Problem # 4	< Same as on left >
Video: Comp. Stoichiometry	< Same as on left >
Study Problem # 5	Worked Ex. of Problem #5
Study Problem # 6	< Same as on left >
Study Problem # 7	Worked Ex. of Problem #7
Study Problem # 8	< Same as on left >
Study Problem # 9	Worked Ex. of Problem #9
Study Problem # 10	< Same as on left >
Post-Questionnaire	< Same as on left >
Video: Introduction to Post Test	< Same as on left >
5 Posttest Problems (Isomorphic to Pretest)	< Same as on left >

Table 2. Populations and Characteristics of the Three Studies

St. #	*N*	Subject Pop.	Notes
1	63	College	o Intro to college chem class o Presented as optional study material o Subjects paid $25 for participation o High drop-out rate, over 100 started o Published in (McLaren *et al*, 2006). After outlier screening, *N* was adjusted from 69 to 63
2	60	High School	o Mix of intro and Advanced Placement ("AP") chem students o Extra credit; very low dropout rate o Briefly cited in (McLaren *et al*, 2007) but otherwise unpublished. After outlier screening, *N* was adjusted from 76 to 60
3	81	High School	o Mix of intro and AP chem students o Extra credit; very low dropout rate o Preliminary results with *N*=33 published in (McLaren *et al*, 2007).

[4] This is the design for studies 2 and 3. There were two differences between study 1 and studies 2 and 3. First, we had to shorten the intervention for use in high schools, the subject population of the latter two studies. There were 9 Pre and Posttest problems and 15 Study Problems in Study 1, instead of 5 and 10, respectively. Second, while there were prompted self-explanation questions after the worked examples in studies 1 and 2, there were none in study 3.

Results

Repeated measure ANOVAs conducted on the pre / posttests in each study revealed significant learning across all conditions (Study 1: $F(1,59)=68.18$, $p<.001$; Study 2: $F(1,56)=77.30$, $p<.001$; Study 3: $F(1,77)=95.71$, $p<.001$). On the other hand, there were no statistically significant main effects in any of the studies due to worked examples, according to ANOVAs done on the difference (post - pre) scores between the Examples and Problems conditions (Study1: $F(1, 61) = 0.005$, n.s.; Study 2: $F(1, 58) = .026$, n.s.; Study 3: $F(1.79) = 1.691$, n.s.). In other words, the subjects in the Examples Condition did not learn more than those in the Problems Condition. These results can be seen visually in the graphs of Figure 2.

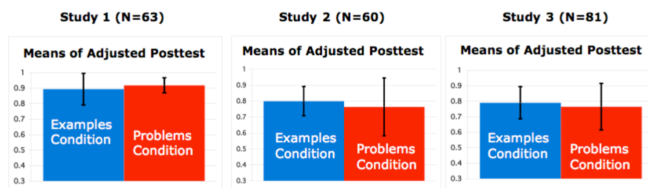

Figure 2. Means of Adjusted Posttests of Studies 1-3

However, subjects in the Examples Condition in all of the studies spent less time with the study problems (of those who did at least ½ of the problems), at a statistically significant level, as shown in Table 3. (This efficiency analysis, as well as the analyses shown in all of the remaining tables, was done after *all* of the studies were completed and thus is first reported here, i.e., these are new results, not reported in (McLaren *et al*, 2006; 2007).)

Table 3**.** Average total time spent doing problems, Examples vs. Problems Conditions. Includes time spent on Study Problems 1 through 10, in Table 1 (1 through 15 for study 1); excludes time spent on pretest, posttest, questionnaires, and videos. The P-value was calculated using ANOVA between the Examples and Problems Conditions' time. Effect size was calculated using Cohen's *d*, with following assumptions: d >= 0.8 (Large effect), d >= 0.5 (Medium effect), d >= 0.2 (Small effect) (Cohen, 1998).

St. #	Examples Condition Avg. Time	Problems Condition Avg. Time	P-P-val.	Effect Size (Cohen's *d*)
1	48 min (sd = 14)	71 min (29)	**0.000***	1.02 (Large)
2	57 min (25)	72 min (25)	**0.029***	0.59 (Medium)
3	64 min (16)	73 min (18)	**0.019***	0.54 (Medium)

* - Significant result

In other words, the subjects in the Examples Condition, while they did not learn more, they learned more efficiently than those in the Problems Condition. This can be seen in Table 4. In studies 1 and 3 the difference between the learning efficiency in the Examples and Problems Conditions was statistically significant in favor of Examples, while in study 2 the difference was not statistically significant but still favored the Examples Condition.

Table 4. Learning Efficiency, calculated, per subject, as z-score (learning gain) - z-score (instructional time) with z-score = (value – average) / standard dev. Values in Table 4 are averages across all subjects. The P-value was calculated using ANOVA between the Examples and Problems Conditions' learning efficiency.

St. #	Examples Condition Learn. Eff.	Problems Condition Learn. Eff.	P-value	Effect Size (Cohen's *d*)
1	0.47	-0.45	**0.005***	0.75 (Medium)
2	0.24	-0.26	**0.146**	0.39 (Small)
3	0.40	-0.41	**0.015***	0.56 (Medium)

* - Significant result

Discussion and Conclusions

In all three of our studies, the results showed that students did not learn more in the alternating Examples Condition, contrary to the findings in earlier studies such as (Trafton & Reiser, 1993; Kalyuga *et al*, 2001). On the other hand, the Examples Condition *did* learn more efficiently, using 21% less time to complete the same problem set. If these results were to scale across a 20-week course, students could save 4 weeks of time – yet learn just as much.

Of course, our studies are different from earlier studies in that they involve tutored problem solving, instead of untutored problem solving. One possible reason for the null learning result is that the students in the Problems Condition equalized themselves to the Examples Condition by using the tutor to create examples through the reading of the bottom-out hints in the tutor (which provide the answer). This might neutralize the expected learning advantage of first studying and then self-explaining examples in the Examples Condition. There is some evidence this occurred, as can be seen in Table 5. In studies 2 and 3 the students in the Problems Condition used the bottom-out hint more when working on the first of the isomorphic example-problem pairs, at a statistically significant level but modest effect size, and in study 1 the comparison was also in this direction, although not significantly so. This provides some support for the hypothesis that students try to make an example out of a tutored problem that is the first of a matched pair of isomorphic problems.

But what explains our finding that the Examples Condition worked more efficiently than the Problems Condition? As can be seen in Table 6, students in the Examples Condition worked much faster on the *first* of the isomorphic example-problem pairs ("Isomorphic Problem n") than the second problem ("Isomorphic Problem n+1"), with a statistically significant interaction effect between the paired problems in the Examples and Problems Conditions in all three studies. In other words, the extra time the students in the Problems Condition take on Isomorphic Problem n – even though it often seems to be used to turn problems into examples, as shown in Table 5 – is not benefiting them. This may be because clicking through hints is a less efficient way to see an example compared to seeing the example immediately, as in the Examples Condition. Or perhaps students in the Problems Condition simply waste more time floundering with the tutor in search of a solution. The difference in time on task between the Examples and Problems conditions cannot be attributed to students skimming the worked examples; we found that students spent, on average, 127% (sd=0.63) of the example video time working on the examples[5].

Table 5. Comparison of bottom-out hints taken per student on the 1st and 2nd problems of the isomorphic pairs in the Problems Condition. The P-value was calculated by a 2-tailed t-test between the number of bottom-out hints in the 1st and 2nd problems across all students. (Note: Statistics were run on all problem pairs except one that was clearly faulty, i.e., one pair of problems was not isomorphic. In this pair, the same terms were required to solve both problems, but in reverse order. Even with this outlier pair included, the difference (and direction) between the Example and Problem conditions was statistically significant in study 2, but not so in studies 1 and 3.)

St. #	Avg. Bottom-Out Hints Isomorphic Problem n	Avg. Bottom-Out Hints Isomorphic Problem n+1	P-val.	Effect Size (Cohen's *d*)
1	3.2 (sd = 6.0)	2.8 (6.0)	**0.320**	0.07 (None)
2	4.1 (5.1)	1.9 (3.0)	**0.002***	0.53 (Med.)
3	5.1 (6.9)	3.1 (5.7)	**0.002***	0.31 (Small)

* - Significant result

Table 6. Comparison of the avg. time spent on the 1st and 2nd problems of the isomorphic pairs in the Examples and Problems Conditions. The int. P-val. was calculated by a 2-way ANOVA.

St #	Condition	Isomorphic Problem n	Isomorphic Problem n+1	P-val.
1	Examples Condition	2.0 min (sd = 1.0)	4.3 min (1.3)	**0.000***
	Problems Condition	4.9 min (1.9)	4.6 min (2.0)	
2	Examples Condition	4.8 min (1.3)	6.7 min (4.5)	**0.001***
	Problems Condition	7.7 min (2.9)	6.6 min (2.4)	
3	Examples Condition	4.8 min (1.5)	5.0 min (1.7)	**0.000***
	Problems Condition	8.2 min (2.9)	4.0 min (1.0)	

* - Significant result

While we did not test for far transfer effects in our studies, prior studies of worked examples and self-explanation have found null effects on normal tests (i.e., near transfer), yet statistically significant effects on far transfer. For example, the study of Schwonke *et al* (2007), similar in many respects to our studies, also got a null effect for normal learning, but a significant effect in favor of the

[5] This includes the time spent video viewing and answering self-explanation questions. The large standard deviation is due to students in study 1 spending only 62% time with the examples. This can be explained by (a) college students being more likely to know the material, thus being more likely to skim, and (b) not being prompted with self-explanation questions as in studies 2 & 3.

Examples Condition, for conceptual transfer. This study illustrates that it is possible the study and self-explanation of examples is more likely to have an effect on *conceptual* learning than on normal learning. The study of Paas and Van Merriënboer (1994) also demonstrated that examples could have a significant effect on transfer learning. While they did not test normal learning – and thus it is unsure they would have gotten null effects – their transfer tests resulted in statistically significant learning gains and efficiency, again in favor of the worked examples condition. We intend to explore this in subsequent studies in which we will include conceptual, transfer questions.

The "minimize cognitive load" theory (Sweller, Van Merriënboer, Paas & 1998) appears to inadequately describe our findings, and we are left with an open theoretical problem. It's possible that all problem solving (or all example study) puts students in a less metacognitive mode – just getting the job done (or just reading the examples), whereas interleaving keeps students more metacognitive by focusing them on (1) reflecting on examples to induce deep regularities (the domain rules), (2) reflecting on whether they got the rule right during problem solving, and (3) returning to the next example more focused on what they don't know yet. That is, they may carry "learning subgoals" from the prior problem into the next example.

Our studies would appear on the dimension of assistance of Figure 1 in like fashion to the Schwonke *et al* studies, in which an all-tutored problems condition was compared to an alternating examples/tutored problems condition (except that our examples have both explained and unexplained portions). Our results are not as strong as theirs with only an efficiency gain in favor of the alternating condition, rather than *both* an efficiency and far transfer gain (i.e., with respect to the key of Figure 1, only a "o" instead of "+o"). Yet our studies are also consistent with the inverted-U hypothesis that mid-level assistance provides the greatest learning advantages, although in less decisive fashion than when the control condition is all untutored problems, as in (cf. Paas, 1992; Trafton & Reiser, 1993). However, we are yet to test the middle range against higher-level assistance (e.g., all worked examples). Thus, our next step in testing the inverted-U hypothesis is to compare three conditions spanning between 2 and 3 on the dimension of assistance of Figure 1: all tutored problems (lower assist.), alternating examples and tutored problems (mid-level assist.), and all unexplained examples (higher assist.).

Acknowledgements. The Pittsburgh Science of Learning Center, NSF Grant # 0354420, supported this research.

References

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.

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

Clark, R.C. & Mayer, R. E. (2003). *E-Learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. Jossey-Bass/Pfeiffer.

Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, A., Brown, J.S., & Newman, S.E. (1990). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser*. Hillsdale, NJ: Lawrence Erlbaum.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Ed. Psych.*, 93, 579-588.

Kirschner, P.A., Sweller, J., & Clark, R.E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86.

Koedinger, K.R. & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*. 19(3), 239-264.

Lovett, M.C. (1992). Learning by problem solving versus by examples: The benefits of generating and receiving information. *Proc. of the 14th Conference of the Cognitive Science Society* (pp. 956-961). Hillsdale, NJ: Erlbaum.

McLaren, B.M., Lim. S., Gagnon, F., Yaron, D., & Koedinger, K.R. (2006). Studying the effects of personalized language and worked examples in the context of a web-based intelligent tutor. *Proc. of the 8th International Conference on Int. Tut. Sys.* (pp. 318-328).

McLaren, B.M., Lim, S., Yaron, D., & Koedinger, K.R. (2007). Can a polite intelligent tutoring system lead to improved learning outside of the lab? *Proc. of the 13th International Conference on AI in Ed.* (pp. 433-440).

Paas, F.G.W.C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Ed. Psych*., 84, 429-434.

Paas, F. & Van Merriënboer, J.J.G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Ed. Psych.*, 86(1), 122-133.

Schwonke, R., Wittwer, J., Aleven, V., Salden, R.J.C.M., Krieg, C., & Renkl, A. (2007). Can tutored problem solving benefit from faded worked-out examples? *Proc. of the 2nd European Cog. Sci. Conference* (pp. 59-64).

Steffe, L. & Gale. J. (Eds.) (1995) *Constructivism in education*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sweller, J., Van Merriënboer, J.J.G., & Paas, F.G.W.C. (1998). Cognitive architecture and instructional design. *Ed. Psych. Review*, 10, 251-296.

Trafton, J.G. & Reiser, B.J. (1993). The contributions of studying examples and solving problems to skill acquisition. *Proc. of the 15th Conference of the Cognitive Science Society* (pp. 1017-1022).

Vygotsky, L.S. (1978). *Mind in society*. Harvard Univ Press.