Exploring the Assistance Dilemma: Comparing Instructional Support in Examples and Problems

Bruce M. McLaren[1], Tamara van Gog[2], Craig Ganoe[1], David Yaron[1], and Michael Karabinos[1]

[1] Carnegie Mellon University, Pittsburgh, PA, USA
{bmclaren,ganoe}@cs.cmu.edu, yaron@cmu.edu, mk7@andrew.cmu.edu
[2] Erasmus University Rotterdam, The Netherlands
vangog@fsw.eur.nl

Abstract. An important question for teachers and developers of instructional software is how much guidance or assistance should be provided to help students learn. This question has been framed within the field of educational technology as the 'assistance dilemma' and has been the subject of a variety of studies. In the study reported in this paper, we explore the learning benefits of four types of computer-based instructional materials, which span from highly assistive (worked examples) to no assistance (conventional problems to solve), with support levels in between these two extremes (tutored problems to solve, erroneous examples). In this never-before conducted comparison of the four instructional materials, we found that worked examples are the most efficient instructional material in terms of time and mental effort spent on the intervention problems, but we did not find that the materials differentially benefitted learners of high and low prior knowledge levels. We conjecture why this somewhat surprising result was found and propose a follow-up study to investigate this issue.

Keywords: assistance dilemma, classroom studies, empirical studies worked examples, erroneous examples, tutored problems to solve, problem solving.

1 Introduction

A major and recurring question for teachers and developers of instructional software is how much assistance they should provide in order to foster students' acquisition of problem-solving skills, i.e., the 'assistance dilemma' [1]. On the high assistance side of the continuum are worked examples, which present students with a fully worked-out problem solution to study and (possibly) explain. On the low assistance side of the continuum are conventional problems, which students try to solve themselves without any instructional guidance whatsoever. In between these two extremes are intelligently-tutored problems, which provide step-by-step feedback and hints either when an error is made or on demand, and erroneous examples, which are worked examples with errors in one or more of the problem-solving steps that students have to find and fix. It is straightforward to place these instructional materials on a continuum of assistance, but an important question is: How can the level and type of assistance best support learners with varying levels of prior knowledge?

S. Trausan-Matu et al. (Eds.): ITS 2014, LNCS 8474, pp. 354–361, 2014.

These types of instructional materials have all been investigated in various empirical studies, in different combinations, although never all together in a single study. For instance, the learning benefits of worked examples have been shown in a plethora of studies (for reviews see [2-4]), particularly for low prior knowledge (i.e., novice) students. Worked examples lessen the demands on cognitive resources, as compared to problem solving, when students are unfamiliar with a problem domain, and allow them to devote available cognitive resources to learning how problems should be solved [4]. In order to foster more active processing of worked examples, successful variations and strategies have been developed [5, 6]. For high prior knowledge learners, worked examples lose their effectiveness or may even become less effective for learning than practicing with conventional problem solving [7], because the assistance provided by the examples is redundant for high prior knowledge learners.

A variety of studies have also demonstrated the learning benefits of intelligently tutored problems [8 9]. Intelligent tutors, like worked examples, tend to benefit lower prior knowledge learners, those who one would expect require the type of support provided by the tutors, more than higher prior knowledge learners [10]. There are also indications that tackling worked examples before working with tutored problems improves learning efficiency (i.e., students learn as much, in less time), and, in some cases, learning outcomes, as compared to tutored problem solving alone [3, 11].

Recent studies – a relatively small number compared to worked examples and intelligently tutored problems – have also investigated the effects of erroneous examples [12-14]. Presenting students with errors might help eradicate those errors by prompting more reflection than would occur naturally. Erroneous examples have so far been shown to be particularly beneficial to learners with some prior knowledge [13], which makes intuitive sense, since a student who has not yet understood the basic concepts and problem-solving procedures within a domain is less likely to be able to differentiate and make sense of correct and incorrect problem solutions.

Finally, as mentioned above, giving students problems to solve, without feedback or support, has been shown to be most beneficial to more advanced students, ones with sufficient prior knowledge to gain from practice without assistance [7].

There is some variability among studies in whether or not feedback was provided to students in the conventional problems group. Paas provided students with feedback on practice problems, which consisted of worked examples. Still, studying worked examples (with a practice problem after two examples) was found to be more effective than practicing with conventional problem solving with feedback [6].

In this study, we intended to compare the learning benefits of these four types of instructional materials (developed for and deployed on the web) at different levels of expertise (lower, higher). Although such comparisons have been partially made, no studies have compared the effectiveness and efficiency of all four support strategies to each other. This study aimed to make that comparison, taking into account students' prior knowledge level in order to take a first step towards testing our hypothesis that worked examples and tutored problem solving are more suitable learning materials for students with lower prior knowledge, while erroneous examples and conventional problem solving are more suitable for students with higher prior knowledge.

2 Method

Participants and Design. Participants were 179 10th and 11th grade students from two high schools in the U.S. Twenty-four participants were excluded because they did not fully complete all phases of the study. The remaining 155 students had a mean age of 15.4 (SD = 0.59), with 75 males, 80 females. Participants were randomly assigned to one of the 4 instructional conditions: (1) Worked Examples (*WE*), (2) Erroneous Examples (*ErrEx*), (3) Tutored Problems to Solve (*TPS*), or (4) Problems to Solve (*PS*).

Materials. A web-based stoichiometry tutor used in earlier studies [3, 15] was revised to support this study. Stoichiometry is a subdomain of chemistry in which basic mathematics (i.e., multiplication of ratios) is applied to chemistry concepts.

Table 1. Conditions and Materials used in the study. *Italicized* items vary across conditions

	WE	*TPS*	*ErrEx*	*PS*
	Questionnaire	Questionnaire	Questionnaire	Questionnaire
	Pretest (A or B)	Pretest (A or B)	Pretest (A or B)	Pretest (A or B)
	WE Intro video	*PS Intro video*	*ErrEx Intro video*	*PS Intro video*
	Stoich videos (both at beginning and interspersed)	Stoich videos (both at beginning and interspersed)	Stoich videos (both at beginning and interspersed)	Stoich videos (both at beginning and interspersed)
x5 {	*WE-1*	*TPS-1*	*ErrEx-1*	*PS-1*
	WE-2	*TPS-2*	*ErrEx-2*	*PS-2*
	Embedded-Test-1	Embedded-Test-1	Embedded-Test-1	Embedded-Test-1
	Posttest (A or B)	Posttest (A or B)	Posttest (A or B)	Posttest (A or B)

Questionnaire. Students were asked demographic, computer use, and self-perceived prior knowledge questions.

Pretest and Posttest. The pretest and posttest consisted of four stoichiometry problems to solve (of the same form as the Intervention Problems) and four conceptual knowledge questions to answer. There was an A and B form of the test, isomorphic to one another and counter-balanced within condition (i.e., approximately ½ of the students in each condition received Test A as pretest and Test B as posttest, and vice versa). The stoichiometry problems consisted of 94 steps in total (one point per correct step). The conceptual questions consisted of 7 possible answers (one point per correct answer). This resulted in a maximum total score of 101 points.

Intro and Instructional videos. After taking the pretest, all students watched a video specific to their condition, which used a narrated example to explain how to interact with the user interface. In addition, students in all conditions were presented with the same instructional videos about stoichiometry and problem solving techniques, starting at the beginning of the intervention and spread throughout the intervention.

Intervention Problems. Students were presented with 10 intervention problems, specific to condition and grouped in isomorphic pairs, as shown in Table 1 (e.g., WE-1 and WE-2 are an isomorphic pair, TPS-1 and TPS-2, etc.). The complexity of the stoichiometry problems presented in the intervention gradually increased.

The *WE* items consisted of problem statements and screen-recorded videos (30-70 sec.) of how to solve the problem. When the video finished, students had to select the "reason" for each step from a drop-down menu. Then they click the "Done" button and feedback appeared. When they were correct, they were encouraged to study the final correct problem state; when they were incorrect a fully worked-out final solution appeared below the problem that students could study self-paced (see Figure 1).

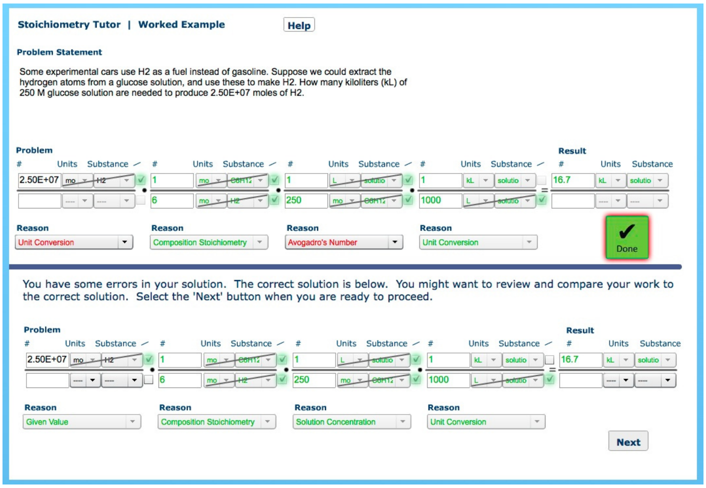

Fig. 1. WE with incorrect reasons resulting in correct worked example feedback

The *TPS* items consisted of a problem statement and fields to fill in (similar to the top of Figure 1) and students had to attempt to solve the problem assisted by on-demand hints and error feedback. There were up to 5 levels of hints per step, with the bottom-out hint giving the answer to that step. Because the tutored problems always ended in a correct final problem state due to the given hints, no feedback was given at the end but students were encouraged to study the final correct problem state.

The *ErrEx* items also consisted of screen-recorded video (30-70 sec.) demonstrating how to solve the problem, except the items contained 1 to 4 errors that students were instructed to find and fix. They had to fix at least one step before they could click 'Done', at which point the same 'correct' or 'incorrect' feedback messages as in the *WE* condition appeared, with a correct example shown if errors were still present.

The *PS* items consisted of a problem statement and fields to fill in (similar to the top of Figure 1) and students had to attempt to solve the problem themselves, without any assistance. They had to fill out at least one step before they could click the 'Done' button. When they clicked the 'Done' button, the same 'correct' or 'incorrect' feedback messages as in the *WE* condition appeared, with a correct example shown if errors were still present.

Embedded test problems. After every two intervention items, an embedded test problem was given that was identical to the first intervention item of the two (i.e., intervention problems 1, 3, 5, 7, and 9), but in *PS* form without any guidance or feedback. These problems consisted of a total of 122 steps (one point per correct step).

Mental effort rating scale. A 9-point mental effort rating scale [6] was administered after each intervention problem.

Procedure. The experiment was conducted at students' schools within their regular science classrooms. In total, the study took 6 class periods to complete. Students received a login for the web-based environment and could work at their own pace (for the phases and tasks they encountered, see Table 1). When they had finished with the intervention phase, however, they could not progress to the posttest; this test took place on the sixth and final period for all students.

3 Results

As mentioned in the introduction, we intended to compare the learning benefits of the four types of instructional materials (developed for and deployed on the web) at different levels of expertise (lower, higher). However, apart from differences in prior knowledge, these analyses did not yield additional insights about the instructional conditions compared to analysis of the overall sample. Because of page limitations, we therefore report only the overall sample results here.

Data are presented in Table 2 and were analyzed with ANOVA. There were no significant differences among conditions in pretest (p = .783)[1], posttest (p = .693), or embedded test problem performance (p = .326).

Table 2. Mean performance, mental effort, and time on task per condition

	***WE* (*n*=39)**	***TPS* (*n*=36)**	***ErrEx* (*n*=43)**	***PS* (*n*=37)**
Pretest (max=101)	48.6 (12.8)	49.4 (13.5)	48.8 (15.4)	46.3 (14.3)
Posttest (max=101)	68.5 (17.3)	71.1 (13.4)	68.3 (18.4)	66.4 (17.1)
Embedded test (max=122)	89.4 (23.7)	95.3 (23.3)	88.3 (27.0)	84.8 (23.1)
Mental eff. inter. probs. (1-9)	4.4 (1.8) *	6.1 (1.7)	5.8 (1.4)	6.1 (1.3)
Intervention time (mins)	19.8 (5.8) *	62.4 (17.2)	37.2 (9.6) #	52.1 (25.2) ~
Reflection time (mins)	1.7 (1.1)	1.3 (1.0)	4.3 (2.6) ^	6.5 (3.9) *

* - Significant difference to all other conditions # - Significant difference to *TPS* and *PS* conditions
^ - Significant difference to *WE* and *TPS* conditions ~ - Significant difference to *TPS*

However, mean mental effort invested on the intervention problems differed significantly among conditions ($p < .001$, $\eta_p^2 = .166$); Bonferroni post hoc tests showed effort was lower in the *WE* condition than in all other conditions (*ErrEx*: $p < .001$, $d = 0.891$; *TPS*: $p < .001$, $d = 0.954$; *PS*: $p < .001$, $d = 1.04$).

Intervention time also differed significantly among conditions ($p < .001$, $\eta_p^2 = .503$); Bonferroni post hoc tests showed that time spent in the *WE* condition was

[1] Due to space limitations, and for readability, only p and effect size values are reported in this paper. F statistics and further statistical details are available from the first author.

lower than in all other conditions (*ErrEx*: $p < .001$, $d = 2.195$; *TPS*: $p < .001$, $d = 3.312$; *PS*: $p < .001$, $d = 1.762$)), in the *ErrEx* condition was lower than in the *TPS* and *PS* conditions (*TPS*: $p < .001$, $d = 1.812$; *PS*: $p < .001$, $d = 0.782$), and in the *PS* condition was lower than in the *TPS* condition ($p = .038$, $d = 0.478$). Note that the last finding makes sense, given that the *TPS* condition also received instructional assistance and feedback *during* intervention problems. Reflection time on the correct worked example given as feedback differed significantly among conditions ($p < .001$, $\eta_p^2 = .418$); Bonferroni post hoc tests showed it was lower in the *WE* and *TPS* conditions (which did not differ from each other, $p = 1.000$) than in the *ErrEx* (*WE* vs. *ErrEx*: $p < .001$, $d = 1.253$; *TPS* vs. *ErrEx*: $p < .001$, $d = 1.507$) and *PS* conditions (*WE* vs. *PS*: $p < .001$, $d = 1.670$; *TPS* vs. *PS*: $p < .001$, $d = 1.848$). Reflection time in the *PS* condition was significantly higher than in all other conditions (*WE* vs. *PS*: $p < .001$, $d = 1.670$; *ErrEx* vs. *PS*: $p < .001$, $d =0.672$; *TPS* vs. *PS*: $p < .001$, $d = 1.848$).

4 Discussion and Conclusions

Our findings suggest that example study was more efficient in terms of the learning process: the *WE* condition attained equal test performance with less time and mental effort on the intervention problems than all other conditions. This is in line with findings from prior studies that compared studying worked examples to conventional problem solving [cf. 16], as well as to tutored problem solving [3, 11].

In contrast to other studies on the worked example effect [6, 7, 16], we did not find a learning outcome benefit for worked examples, either overall or in the lower prior knowledge sample. Also, our hypothesis that the instructional materials would be differentially beneficial to learners based on prior knowledge level was not supported. A distinguishing aspect of this study is the use of a common user interface for conditions ranging from the highly assistive (*WE*) through unassisted problem solving (*PS*). In *WE* and *ErrEx*, the examples are implemented as videos of problem solving within the interface. In PS and TPS, students use the interface to solve problems, with conditions differing with regard to immediate versus delayed feedback. This design has the advantage of allowing tight control of conditions, with differences arising only in the nature of student interaction with the interface. The finding of equal learning gains across conditions is interesting, given the substantial differences in the nature of the student interactions as well as in the mental effort and time across condition.

A common feature across conditions that may account for these findings is the presence of a fully and correctly worked example at the end of each problem-solving episode, which students could study as long as they wished. We provided students with feedback in order to make the comparison among the conditions as fair as possible; however, providing feedback in the form of fully worked-out solutions led to a very strong presence of worked examples in *every* condition. *TPS* students generate the solution, but they also effectively get worked examples by drilling down to bottom-out hints. In the *ErrEx* and *PS* conditions, in which errors occurred often (81% of the time) and a correct example was then provided, the mean time spent reflecting on comparing student work to a correctly worked example (*ErrEx* = 31.1 secs and

PS = 42.8 secs) is comparable to the amount of time students in the *WE* condition spent watching the animated worked example (i.e., between 30 and 70 seconds, as earlier mentioned). Few other studies [cf. 6] on the worked example effect provided students in the *PS* condition with worked examples as feedback, and in those studies they could review the feedback for a restricted amount of time that was less than the amount of time students in our *WE* condition could study the examples.

Because the use of worked examples may have made the conditions too similar, we will next run a study in which the conditions will be more distinct. We will drop the worked examples as a form of feedback in the *WE*, *ErrEx*, and *PS* conditions. Instead of receiving the correct worked example as feedback, students will only see feedback highlighting the steps they correctly and incorrectly complete.

Acknowledgement. National Science Foundation Award No SBE-0354420 ("Pittsburgh Science of Learning Center") funded this research.

References

1. Koedinger, K.R., Aleven, V.: Exploring the assistance dilemma in experiments with cognitive tutors. Educational Psychology Review 19, 239–264 (2007)
2. Atkinson, R.K., Derry, S.J., Renkl, A., Wortham, D.: Learning from examples: Instructional principles from the worked examples research. Review of Ed R'ch 70, 181–214 (2000)
3. 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: Love, B.C., McRae, K., Sloutsky, V.M. (eds.) Proceedings of the 30th Annual Conference of the Cog. Sci. Society, pp. 2176–2181. Cognitive Science Society, Austin (2008)
4. Sweller, J., Van Merriënboer, J.J.G., Paas, F.: Cognitive architecture and instructional design. Educational Psychology Review 10, 251–295 (1998)
5. 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. CogSci. 13, 145–182 (1989)
6. Paas, F.: Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. Journal of Educational Psychology 84, 429–434 (1992)
7. 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)
8. VanLehn, K.: The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. Educational Psychologist 46(4), 197–221 (2011)
9. Graesser, A.C., Chipman, P., Haynes, B.C., Olney, A.: AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. IEEE Transactions in Ed. 48, 612–618 (2005)
10. Ritter, S., Anderson, J.R., Koedinger, K., Corbett, A.: Cognitive Tutor: Applied research in mathematics education. Psychonomic Bulletin and Review 14, 249–255 (2007)
11. Salden, R., Koedinger, K.R., Renkl, A., Aleven, V., McLaren, B.: Accounting for beneficial effects of worked examples in tutored problem solving. Ed. Psy. Rev. 22(4), 379–392 (2010)
12. Durkin, K., Rittle-Johnson, B.: The effectiveness of using incorrect examples to support learning about decimal magnitude. Learning and Instruction 22, 206–214 (2012)

13. Grosse, C.S., Renkl, A.: Finding and fixing errors in worked examples: Can this foster learning outcomes? Learning and Instruction 17, 612–634 (2007)
14. McLaren, B.M., et al.: To err is human, to explain and correct is divine: A study of interactive erroneous examples with middle school math students. In: Ravenscroft, A., Lindstaedt, S., Kloos, C.D., Hernández-Leo, D. (eds.) EC-TEL 2012. LNCS, vol. 7563, pp. 222–235. Springer, Heidelberg (2012)
15. McLaren, B.M., DeLeeuw, K.E., Mayer, R.E.: Polite web-based intelligent tutors: Can they improve learning in classrooms? Computers & Education 56(3), 574–584 (2011)
16. Van Gog, T., Paas, F., Van Merriënboer, J.J.G.: Effects of process-oriented worked examples on troubleshooting transfer performance. Learning and Instruction 16, 154–164 (2006)