

Worked Examples are More Efficient for Learning than High-Assistance Instructional Software

Bruce M. McLaren¹(✉), Tamara van Gog², Craig Ganoe¹, David Yaron¹,
and Michael Karabinos¹

¹ Carnegie Mellon University, Pittsburgh, PA, USA
bmclaren@cs.cmu.edu, ganoe@acm.org, yaron@cmu.edu,
mk7@andrew.cmu.edu

² Erasmus University Rotterdam, Rotterdam, The Netherlands
vangog@fsw.eur.nl

Abstract. The ‘assistance dilemma’, an important issue in the Learning Sciences, is concerned with how much guidance or assistance should be provided to help students learn. A recent study comparing three high-assistance approaches (worked examples, tutored problems, and erroneous examples) and one low-assistance (conventional problems) approach, in a multi-session classroom experiment, showed equal learning outcomes, with worked examples being much more efficient. To rule out that the surprising lack of differences in learning outcomes was due to too much feedback across the conditions, the present follow-up experiment was conducted, in which feedback was curtailed. Yet the results in the new experiment were the same: there were no differences in learning outcomes, but worked examples were much more efficient. These two experiments suggest that there are efficiency benefits of worked example study. Yet, questions remain. For instance, why didn’t high instructional assistance benefit learning outcomes and would these results hold up in other domains?

Keywords: Assistance dilemma · Classroom studies · Empirical studies · Worked examples · Erroneous examples · Tutored problems to solve · Problem solving

1 Introduction

An important question for the Learning Sciences to answer is how much guidance or assistance should be provided in order to help students learn, i.e., the ‘assistance dilemma’ [1]. In a recent experiment [2], the effectiveness and efficiency of three high-assistance instructional formats (which differ in the amount of student activity required) was compared to low-assistance problems to solve, which students have to attempt to solve largely on their own:

- *worked examples*, which present students with a fully worked-out problem solution to study;
- *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 was found that worked example study resulted in a large efficiency benefit compared to all other conditions. Equal learning outcomes were attained in 50-65% less time and with less self-reported effort invested in the study phase [2]. That the more passive high-assistance format was most efficient is interesting in light of the assistance dilemma. However, it was remarkable that none of the high-assistance instructional formats improved learning outcomes compared to problem solving. Possibly, the feedback received in all of the conditions, including problem solving, in the form of a worked example if they made mistakes, could explain the lack of effect.

To find out whether worked example feedback contributed to the equal performance across conditions, a second study was conducted, reported here. Instead of receiving a correct worked example as feedback, students in all conditions would instead see highlighting of their correct steps in green, and their incorrect steps in red. Thus, the second study, like the earlier one, also directly compared the four instructional conditions, but under different (and lesser) feedback circumstances.

2 Method

Participants and Design. Participants were 116 tenth and eleventh grade students from two high schools in the U.S. ($M_{\text{age}} = 16.45$, $SD = 0.76$; 48 male; 15 of an original 131 participants were excluded for not fully completing all phases). Participants were randomly assigned to one of the four instructional conditions: (1) Worked Examples (*WE*), (2) Erroneous Examples (*ErrEx*), (3) Tutored Problems to Solve (*TPS*), or (4) Problems to Solve (*PS*).

Materials and Procedure. We used the same web-based stoichiometry-learning environment as [2]. Stoichiometry is a subdomain of chemistry in which basic mathematics (i.e., multiplication of ratios) is applied to chemical quantities such as mass and solution concentration. The experiment was conducted at students' schools as replacement for their regular science class. 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 on the materials they encountered in the learning phase.

They first completed a demographic questionnaire, followed by the pretest, consisting of four stoichiometry problems to solve (isomorphic to the Intervention Problems, described below) and four conceptual knowledge questions (max. score: 101 points). Subsequently, each condition watched an introductory video explaining how to interact with the web-based user interface. They then watched instructional videos introducing new stoichiometry concepts and procedures (the same in all conditions), after which students were presented with a total of 10 intervention problems, in an instructional format specific to their condition (explained below). The problems were grouped in five isomorphic pairs (e.g., WE-1 and WE-2 are an isomorphic pair, WE-3 and WE-4 are an isomorphic pair, etc.) and each pair was followed by an isomorphic embedded test problem (max. total score: 122 points). After each intervention item, students indicated how much mental effort they invested in studying/completing it, on a 9-point rating scale [3]. The complexity of the stoichiometry problems gradually increased, with each pair of intervention problems being preceded by instructional videos explaining new concepts and procedures. When they had finished with the

intervention phase, however, they could not immediately progress to the posttest; this test took place on the sixth and final period for all students and was isomorphic to the pretest (max. score: 101 points). Performance was automatically scored, along with time on task and student self-reports of effort.

The worked examples (*WE*) consisted of problem statements and screen-recorded videos of the solution to the problem being entered, step-by-step, into the interface used in all four conditions. The videos had duration of between 30 and 70 seconds, and did not include any narration or explanation of why steps were taken; students only saw the steps being completed. When the video finished, students had to indicate the “reason” for each step by selecting this from a drop-down menu. After entering reasons, they could click the “Done” button and correct/incorrect feedback appeared (in the form of green highlighting for correct steps, red highlighting for incorrect steps). The Erroneous Examples (*ErrEx*), consisted of screen-recorded videos of 30 to 70 seconds, except the items contained 1 to 4 errors that students were instructed to find and fix. They had to correct at least one step before they could click the ‘Done’ button, at which point correct/incorrect feedback appeared. The tutored problems (*TPS*) consisted of a problem statement and fields to fill in and students had to attempt to solve the problem themselves, but with assistance received in the form of on-demand hints and error feedback. There were up to 5 levels of hints per step, with the bottom-out hint being both a message giving the answer to that step and a worked example of the problem solved to that point, shown below the interface. Because the tutored problems always ended in a correct final problem state, students received no further feedback. The problems to solve (*PS*) consisted of a problem statement and fields to fill in by students 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, correct/incorrect feedback appeared. In all conditions, students could review their work – including correct/incorrect feedback – for as long as they wanted before selecting a “Next” button and proceeding to the next item.

3 Results

Data are presented in Table 1 and were analyzed with ANOVA. Analysis of the pretest scores confirmed that there were no significant differences among conditions in prior knowledge, $F(3,112) < 1$, $p = .500$. Test performance did not differ significantly among conditions, either on the embedded test problems, $F(3,112) = 1.031$, $p = .382$, or on the posttest, $F(3,112) < 1$, $p = .883$.

There was a significant difference among conditions in the average mental effort invested in the intervention problems, $F(3,112) = 9.709$, $p < .001$; Bonferroni post-hoc tests showed that $WE < TPS$ ($p < .001$) and PS ($p = .002$); $ErrEx < TPS$ ($p = .003$); no other differences were significant. Regarding time spent on the intervention problems, significant differences among conditions were found $F(3,112) = 72.93$, $p < .001$. Bonferroni post-hoc tests showed: $WE < \text{than all others}$, all $p < .001$; $ErrEx < \text{than TPS and PS}$, both $p < .001$; $TPS > \text{than PS}$, $p = .014$.

Table 1. Performance, mental effort, and time on task per condition. **Sig diffs indicated by ***

	WE (<i>n</i> = 29)	ErrEx (<i>n</i> = 28)	TPS (<i>n</i> = 27)	PS (<i>n</i> = 32)
Pre-test (0-101)	48.69 (17.62)	47.46 (20.27)	41.85 (16.77)	45.25 (16.26)
Embedded test (0-122)	92.21 (25.03)	79.79 (33.29)	85.30 (30.94)	80.69 (31.08)
Effort intervention (1-9)	4.88 (1.44)*	5.31 (1.71)	6.70 (1.25)	6.27 (1.31)
Time intervention (min.)	20.87 (5.50)*	40.48 (11.27)	67.11 (18.91)	56.82 (11.79)
Post-test (0-101)	68.21 (18.21)	65.68 (23.08)	67.78 (19.98)	69.88 (19.17)

4 Discussion and Conclusions

This study replicated the findings of [2], so across the two experiments, evidence was found for enormous efficiency benefits of worked example study, both in terms of effort and time investment, compared to all other conditions (except for effort on the erroneous examples in the present experiment).

The high-assistance instructional formats did not result in better learning outcomes than problem solving. We can only speculate about potential causes. One possibility is that the instructional videos on stoichiometry that were interspersed throughout the intervention in all conditions, and which sometimes included an example of how to apply a concept during problem solving, provided sufficient support for students in the problem-solving condition to benefit from practice, although that was slower and more effortful.

These surprising, and now replicated, results are worthy of further study, especially given that the interspersed conceptual videos, providing both theoretical and procedural explanations, are much closer to real educational practice and therefore give more ecologically valid information about the impact of various instructional formats on learning processes and outcomes.

This study, conducted in a classroom context, finds a clear time-efficiency advantage to worked examples. This result is a valuable finding for educational practice, although one that should be verified in additional domains, with different materials.

Acknowledgements. The National Science Foundation funded this research, Award No. SBE-0836012 (“Pittsburgh Science of Learning Center”).

References

1. Koedinger, K.R., Aleven, V.: Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review* **19**, 239–264 (2007)
2. McLaren, B.M., van Gog, T., Ganoë, C., Yaron, D., Karabinos, M.: Exploring the assistance dilemma: comparing instructional support in examples and problems. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) *ITS 2014*. LNCS, vol. 8474, pp. 354–361. Springer, Heidelberg (2014)
3. 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)