CORNELIA S. GROSSE & ALEXANDER RENKL

LEARNING FROM WORKED EXAMPLES: WHAT HAPPENS IF ERRORS ARE INCLUDED?

Abstract. Learning from worked examples is a very effective learning method. But what happens if errors are included? In two experiments it was explored if a combination of correct and incorrect worked examples enhances learning outcomes, and if incorrect examples are more effective when the error is highlighted. In addition, the effectiveness of fostering self-explanations was assessed. In Experiment 1, the participants learned to solve probability problems under six conditions that constituted a 2x3-factorial design (Factor "incorrect solutions": incorrect solutions with highlighting the error vs. incorrect solutions without highlighting the error vs. correct solutions; Factor "self-explanations": prompting written self-explanations vs. no prompting). It was found that incorrect solutions can foster far transfer performance if learners have favorable domain-specific learning prerequisites. Experiment 2 replicated this effect. Thus, incorrect examples can enhance learning outcomes, but only for "good" learners.

1. INTRODUCTION

Learning from worked examples is an effective learning method in well-structured domains such as mathematics. In this article, a new feature for the design of worked examples is introduced: Learning with incorrect examples. At first glance it might seem surprising to show learners errors, when the goal is to learn to solve problems correctly. Why does it make sense to explore the potential of incorrect examples?

International studies have shown that the mathematical knowledge of students in Western countries such as Germany and the U.S. is quite poor compared to that of Japanese students (e.g., Baumert et al., 2001). In attempts to explain these differences it was argued that Japanese teachers place a particular focus on *how to find* the solution. Incorrect solutions are discussed when presented by students to the class, and the other students try to locate and correct the errors (Klieme & Baumert, 2001). Another argument comes from models which emphasize the relevance of errors in a broader sense (i.e., errors in a solution method or in understanding). For example, VanLehn (1999) argues within his CASCADE model that errors can enhance reflection, which in turn leads to deeper understanding *(impasse-driven learning)*.

2. LEARNING FROM WORKED EXAMPLES

A worked example consists of a problem, the solution steps and the final solution. In comparison to learning by problem solving, learning from worked examples is very effective (e.g., Sweller & Cooper, 1985), which can be explained by the Cognitive Load Theory (e.g., Paas, Renkl, & Sweller, 2003). When solving problems most cognitive resources are required to actually solve the problem, which is usually accomplished by the resource-demanding means-ends-analysis strategy. Thus, there remains only little capacity for learning processes (e.g., construction of abstract schemata). On the other hand, when studying worked examples, the learner is freed from performance demands and can concentrate on gaining understanding.

For effective learning with worked examples, the activity of the learner is crucial. Chi, Bassok, Lewis, Reimann, and Glaser (1989) found that learners who try to explain the examples to themselves learn more. Renkl (1997) has shown that most learners are passive or superficial self-explainers, which emphasizes the importance of prompting self-explanations. In recent years, prompts of written self-explanation have been taken into consideration. For example, Hausmann and Chi (2002) as well as Schworm and Renkl (2002) analyzed the effects of such prompts and found that focused prompting substantially fostered learning outcomes.

Furthermore, learning outcomes depend on the design of the examples. In this context, *intra-example features* (features of single examples) and *inter-example features* (characteristics of combinations of worked examples and problems to be solved) can be distinguished (Atkinson, Derry, Renkl, & Wortham, 2000). One of these features, called a *structure-emphasizing example set* (Quilici & Mayer, 1996), is important in the present context. Structure-emphasizing example sets are relevant when the differentiation between problem types has to be learned. The problem of mixing up different problem types is partly caused by the tendency to judge the similarity of problem types not by their underlying structure, but by random surface features. In order to help learners to discriminate the underlying structure the following procedure is helpful: In different examples which share the same underlying structure, different surface stories are implemented. The same surface stories are then used with other problem types. Thus, it is demonstrated to the learners that while using the same surface story, a problem may have a different underlying structure. In this way students can learn to categorize problems according to structural aspects. In the present experiments, we test whether there is added value from including incorrect examples in a structure-emphasizing example set.

3. LEARNING WITH INCORRECT SOLUTIONS

It is probably not very useful to present learners solely examples containing errors. Learning with incorrect solutions means that after the initial presentation of correct examples, examples with errors are presented in order to deepen the already acquired knowledge. The kind of the error should be adapted to the learning goal. For example, when errors are implemented which confuse two structures, the learners can self-explain the error by referring to the structure of the given task. Then, they can self-explain why the presented solution is wrong and how it could be corrected. It is also possible to think about circumstances under which the presented solution would be correct. It is an open question to what extent the learners spontaneously provide such error-triggered self-explanations. Self-explanation prompts might be necessary to render learning from incorrect examples effective.

Learning with incorrect examples poses challenging demands on the learners. They have to represent not only the correct solution in their working memory, but also the incorrect step with an explanation why it is wrong. Learners with low prior knowledge who cannot form larger chunks for information coding can easily be overtaxed. Thus, from a Cognitive Load Theory perspective the advantage of learning with worked examples should disappear in the case of a combination of incorrect examples and low prior knowledge. An instructional possibility to lower the demands is an explicit specification of the error, so that the learners do not have to search for it on their own. For learners with low prior knowledge, this could reduce cognitive load and foster learning outcomes. On the other hand, for learners with high prior knowledge, this procedure could produce negative effects because the incentive to think carefully about the whole solution is reduced.

4. EXPERIMENT 1

4.1. Research Questions

The following research questions were addressed: (1) Does the presentation of incorrect solutions foster learning outcomes? (2) Does prompting self-explanations foster learning outcomes? (3) Does the effectiveness of incorrect solutions depend on prior knowledge, prompting self-explanations, or highlighting errors?

4.2. Methods

4.2.1. Sample and Design

The participants were 118 students of the University of Freiburg (Germany). A 3x2-design was implemented (see Table 1). The factor "incorrect solutions" concerns the presence or absence of incorrect solution steps, the factor "self-explanation prompts" concerns the presence or absence of prompts for written self-explanations.

Table 1. Design and sample size

	Correct solutions	*Incorrect solutions without highlighting the error*	*Incorrect solutions + highlighting the error*
Without self-explanation prompts	$n = 20$	$n = 20$	$n = 20$
With self-explanation prompts	$n = 20$	$n = 19$	$n = 19$

4.2.2. Materials

The learning materials contained 8 examples from the domain of probability. The examples 1-4 demonstrated the easily mixed-up structures "relevant vs. irrelevant sequence". To foster the differentiation of those structures, two examples with different surface stories were presented first, demonstrating different structures (see Table 2, columns 1 and 2). Then, two examples with the same structures but with interchanged surface stories were presented (columns 3 and 4). Thus, the "correct" groups constitute "very strong" control groups with structure-emphasizing example sets. In the "incorrect" groups, the same two correct examples as in the "correct" groups were presented first (columns 1 and 2). Then, two examples were presented whose solutions showed a mix-up error (columns 3 and 4). The learners were informed that the presented solutions were not correct. In the "highlighting" groups, the specific error was marked. The examples 5-8 demonstrated the structures "with vs. without intersection" and were designed accordingly.

Table 2. Structure of examples 1-4

Example 1 structure "irrelevant sequence" surface story A	Example 2 structure "relevant sequence" surface story B	Example 3 structure "irrelevant sequence" surface story B	Example 4 structure "relevant sequence" surface story A
	Time		
For all groups: correct solution: "irrelevant sequence"	*For all groups:* correct solution: "relevant sequence"	*For groups with correct solutions:* "irrelevant sequence"	*For groups with correct solutions:* "relevant sequence"
		For groups with incorrect solutions: "relevant sequence"	*For groups with incorrect solutions:* "irrelevant sequence"

In the groups with self-explanation prompts, the learners were asked to answer questions in a written format. In order to assure ecological validity the prompts differed across the groups. In the group "incorrect solutions without highlighting the error" the prompts were (1) "Which step is not correct, and why?" (2) "What would be the correct solution?" (3) "Can you give a problem for which the presented solution would be correct?" In the group "incorrect solutions with highlighting the error" the first prompt was changed to "Why is the indicated step not correct?" The second and third prompts remained the same. For the "correct solutions group", the prompts were (1) "Which solution method is used?" (2) "Why is this method adequate?" (3) "Where does one have to pay special attention in order to avoid a mistake?"

4.2.3. Procedure and Instruments

The participants first worked on a pretest containing 11 probability problems in order to assess prior topic knowledge. To control for prior domain knowledge in mathematics, the participants were asked to specify their last school grade in mathematics (in Germany 1 indicates very good performance and 6 indicates very poor performance). Then, a short instructional text was presented which covered the main aspects necessary to understand the examples. Afterwards, 8 examples were presented; here the experimental variation took place. Finally, the participants worked on a post-test which assessed near transfer (4 problems with the same structures but different surface stories as the examples), medium transfer (4 problems with different structures but the same surface stories as the examples), and far transfer (7 problems with different structures and different surface stories). Each correctly solved problem in the pretest and post-test was awarded 1 point.

4.3. Results and Discussion

Concerning the pretest and mathematics grade, no significant group differences were found (all $p > .10$). Thus, the participants in all conditions had about the same prior topic knowledge and domain knowledge (see Table 3).

Table 3. Means (and standard deviations) of prior knowledge and learning outcome in the experimental groups

		Pretest	*Math. grade*	*Near transfer*	*Medium transfer*	*Far transfer*
Correct solutions	*+ SE*	7.68 (2.43)	2.22 (1.21)	3.00 (.96)	2.78 (1.08)	4.47 (1.75)
	No SE	8.10 (2.37)	1.74 (.89)	2.95 (1.37)	3.00 (1.12)	4.66 (1.73)
Incorrect solutions + highlighting error	*+ SE*	8.18 (2.17)	2.29 (.92)	3.03 (.86)	2.58 (1.20)	4.18 (1.93)
	No SE	7.93 (2.12)	1.93 (.88)	3.00 (.71)	2.80 (1.13)	4.87 (1.84)
Incorrect solutions without highlighting error	*+ SE*	8.47 (1.86)	1.75 (.79)	2.92 (1.23)	2.92 (1.17)	4.25 (2.00)
	No SE	8.23 (2.18)	1.83 (1.23)	2.83 (1.29)	3.10 (1.12)	4.43 (2.30)

Note. SE = self-explanation prompts.

In order to analyze the effects of the experimental factors and the interaction effects with prior knowledge in probability or in mathematics, ANOVAs with completely specified designs were computed. For near transfer, an analysis with prior topic knowledge in probability calculation only revealed a significant influence of the pretest ($F(1, 110) = 27.42$, $p < .001$, $\eta^2 = .200$, strong effect; all other effects: $p > .10$). An analysis with the mathematics grade as an indicator for prior domain knowledge only revealed a significant influence of the mathematics grade ($F(1, 108) = 12.62$, $p = .001$, $\eta^2 = .105$, medium to strong effect; all other effects: $p > .10$). Thus, neither incorrect solutions nor self-explanation prompts enhanced near transfer. For medium transfer, the only significant results were found for the pretest ($F(1, 110) = 49.31$, $p < .001$, $\eta^2 = .310$, strong effect) and for the mathematics grade ($F(1, 108) = 7.91$, $p = .006$, $\eta^2 = .068$, medium effect; all other effects: $p > .10$). Thus, neither incorrect solutions nor self-explanation prompts were suitable in enhancing medium transfer. In the case of far transfer, differences between the experimental groups emerged. Whereas the inclusion of the pretest did not lead to significant differences ("pretest": $F(1, 110) = 66.78$, $p < .001$, $\eta^2 = .378$, strong effect; all other effects: $p > .10$), taking the mathematics grade into consideration led to significant differences between conditions ("incorrect solutions": $F(1, 108) = 4.85$, $p = .030$, $\eta^2 = .043$, small to medium effect; "mathematics grade": $F(1, 108) = 24.21$, $p < .001$, $\eta^2 = .183$, strong effect; interaction: $F(1, 108) = 6.63$, $p = .011$, $\eta^2 = .058$, medium effect; all other effects: $p > .10$). This indicates that learning with incorrect examples led – in general – to worse learning outcomes in far transfer than learning with correct examples ("incorrect solutions": $M = 4.44$, $SD = 2.00$; "correct solutions": $M = 4.56$, $SD = 1.72$). However, the significant interaction with the prior domain knowledge indicates that for learners with favorable prerequisites, a combination of correct and incorrect examples led to better performance in far transfer than a pure presentation of correct examples.

Does highlighting the error enhance learning outcomes? The two groups "incorrect solutions with highlighting the error" (with and without self-explanation prompts) were compared with the two groups "incorrect solutions without highlighting the error" (with and without self-explanation prompts). Analyses including the pretest did not lead to significant results with respect to the experimental conditions (near transfer: "pretest": $F(1, 74) = 32.42$, $p < .001$, $\eta^2 = .305$, strong effect; all other effects: $p > .10$; medium transfer: "pretest": $F(1, 74) = 36.32$, $p < .001$, $\eta^2 = .329$, strong effect; all other effects: $p > .10$; far transfer: "pretest": $F(1, 74) = 48.82$, $p < .001$, $\eta^2 = .397$, strong effect; all other effects: $p > .10$). However, including the mathematics grade in the analyses led to an interesting result. Whereas in the case of near and medium transfer no significant effects of the experimental conditions were observed (near transfer: "mathematics grade": $F(1, 74) = 13.10$, $p = .001$, $\eta^2 = .150$, strong effect; all other effects: $p > .10$; medium transfer: "mathematics grade": $F(1, 74) = 13.31$, $p < .001$, $\eta^2 = .152$, strong effect; all other effects: $p > .10$), in the case of far transfer a tendentially significant difference emerged ("highlighting the error": $F(1, 74) = 3.46$, $p = .067$, $\eta^2 = .045$, small to medium effect; "mathematics grade": $F(1, 74) = 52.93$, $p < .001$, $\eta^2 = .417$, strong effect; interaction: $p > .10$). In the far transfer problems, the learners who were shown the specific error achieved better results than the learners who had to find the error on their own ("incorrect solutions with highlighting the error": $M = 4.53$, $SD = 1.89$; "incorrect solutions without highlighting the error": $M = 4.34$, $SD = 2.13$). Thus, highlighting the error in incorrect solutions can foster far transfer performance.

Experiment 1 showed that in order to enhance near or medium transfer, the presentation of correct and incorrect solutions was not more effective than a structure-emphasizing example set with correct solutions only. In the case of far transfer, incorrect solutions can lead to enhanced learning outcomes if the learner has favorable mathematical domain knowledge. Highlighting the specific error can foster far transfer. A second experiment was conducted in order to replicate these findings.

5. EXPERIMENT 2

5.1. Research Questions

The following research questions were asked: (1) Does the presentation of correct and incorrect solutions foster learning outcomes (at least far transfer) compared to the presentation of correct solutions only? (2) Is this only true if learners have good domain knowledge?

5.2. Methods

Forty students of the University of Freiburg (Germany) were divided equally into two experimental groups. The participants of one group learned with correct and incorrect examples ("incorrect solutions"; error was not highlighted), the participants of the second group learned with correct examples only ("correct solutions"). Materials and procedure were the same as in Experiment 1.

5.3. Results and Discussion

Concerning the pretest and mathematics grade, no significant group differences were found ($p > .10$). Thus, the participants in both conditions had the same prior topic and domain knowledge (see Table 4).

Table 4. Means (and standard deviations) of prior knowledge and learning outcome in the experimental groups

	Pretest	*Mathematics grade*	*Near transfer*	*Medium transfer*	*Far transfer*
Correct solutions	9.53 (2.15)	2.09 (1.19)	2.98 (.83)	3.03 (.91)	4.48 (1.67)
Incorrect solutions	8.88 (2.57)	2.33 (1.16)	2.83 (1.04)	2.96 (1.00)	4.35 (1.82)

For near transfer, no significant group differences were found (analysis with pretest: "pretest": $F(1, 36) = 12.76$, $p = .001$, $\eta^2 = .262$, strong effect; all other effects: $p > .10$; analysis with mathematics grade: all $p > .10$). Thus, the presentation of incorrect solutions did not affect near transfer, and no interaction with prior topic or domain knowledge was observed. Whereas prior topic knowledge significantly influenced the learning outcomes, no effect was found for domain knowledge. For medium transfer, an analysis with the pretest did not yield significant group differences ("pretest": $F(1, 36) = 9.22$, $p = .004$, $\eta^2 = .204$, strong effect; all other effects: $p > .10$); thus, incorrect solutions did not enhance medium transfer. However, an analysis with the mathematics grade yielded a tendentially significant interaction ($F(1, 36) = 3.13$, $p = .085$, $\eta^2 = .080$, medium to strong effect; all other effects: $p > .10$). This indicates that for learners with good domain knowledge, learning with incorrect solutions enhances medium transfer, whereas for learners with poor domain knowledge, learning with correct solutions is more effective. For far transfer, an analysis with consideration of topic knowledge did not lead to significant differences ("pretest": $F(1, 36) = 10.27$, $p = .003$, $\eta^2 = .222$, strong effect; all other effects: $p > .10$). However, including domain knowledge in the analysis yielded a tendentially significant interaction ($F(1, 36) = 2.89$, $p = .098$, $\eta^2 = .074$, medium to strong effect; all other effects: $p > .10$). Participants with good domain knowledge profited more from incorrect solutions, whereas participants with poor domain knowledge learned more from correct solutions. To sum up, it can be stated that the results of Experiment 1 were confirmed. Learning with incorrect solutions can enhance transfer when the learner has favorable prior domain knowledge.

6. GENERAL DISCUSSION

Both experiments showed that for learners with favorable prior domain knowledge, learning with incorrect solutions can enhance transfer. Learning with incorrect solutions is not recommended for learners who have problems in understanding mathematics. Good topic knowledge does not help when learning with incorrect solutions, but sound mathematical knowledge seems to be a prerequisite for effective learning with incorrect solutions. Obviously a good mathematical background is necessary to enable the learner to detect and correct errors, even if little topic knowledge is available. Concerning instructional design it can be concluded that in early stages of the learning process only correct solutions should be provided, whereas in later stages the presentation of incorrect examples can be effective.

Self-explanation prompts did not foster learning outcomes, but it is possible that unfavorable self-explanations were prompted. Further experiments are necessary to explore which types of self-explanations are especially effective when learning with incorrect solutions.

In both experiments, the learners in the "incorrect solutions" groups were not given the correct solutions, so they did not receive feedback to their own elaborations about how to correct the error. In further experiments it could be tested to what extent feedback enhances learning outcomes. Furthermore, it is possible to implement this learning method in computer-based learning environments where it is possible to link incorrect solutions with correct ones and enable the learners to compare their own solutions with correct ones.

AFFILIATIONS

Cornelia S. Grosse and Alexander Renkl, University of Freiburg, Department of Psychology, Engelbergerstr. 41, 79085 Freiburg, Germany

REFERENCES

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. W. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70,* 181-214.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J., & Weiß, M. (Hrsg.). (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.

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.

Hausmann, R. G. M., & Chi, M. T. H. (2002). Can a computer interface support self-explaining? *Cognitive Technology, 7,* 4-14.

Klieme, E., & Baumert, J. (Hrsg.). (2001). *TIMSS – Impulse für Schule und Unterricht. Forschungsbefunde, Reforminitiativen, Praxisberichte und Video-Dokumente.* Bonn: Bundesministerium für Bildung und Forschung.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38,* 1-4.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology, 88*, 144-161.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21,* 1-29.

Schworm, S. & Renkl, A. (2002). Learning by solved example problems: Instructional explanations reduce self-explanation activity. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society* (pp. 816-821). Mahwah, NJ: Erlbaum.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2,* 59-89.

VanLehn, K. (1999). Rule-learning events in the acquisition of a complex skill: An evaluation of CASCADE. *The Journal of the Learning Sciences, 8,* 71-125.