Learning From Errors – The Effect of Comparison Prompts in Instruction After Problem Solving Settings

Katharina Loibl, University of Education Freiburg, katharina.loibl@ph-freiburg.de
Timo Leuders, University of Education Freiburg, leuders@ph-freiburg.de

Abstract: Students, who engage in problem-solving activities targeting yet to-be-learned concepts, usually generate erroneous or incomplete solution attempts. These erroneous solution attempts can form the basis for acquiring valid target concepts during subsequent instruction. Literature on conceptual change as well as studies on 'productive failure' indicate that elaborating on typical errors and comparing these erroneous solution attempts to correct solutions may be crucial for learning in these settings. We compared three conditions in an experimental study: Students of all conditions first engaged in an identical problem-solving activity. Afterwards students worked on elaboration tasks that introduced correct solutions. In this so-called instruction phase, students worked with 1) only correct solutions, 2) correct and typical erroneous solution attempts, 3) correct and typical erroneous solution attempts with prompts to compare these attempts. Posttest results indicate that only students who were prompted to compare the solution attempts significantly benefited from learning with erroneous solution attempts.

Introduction

Imagine the following scenario: students attempt to solve a problem to which they have not yet learnt a solution in school. While struggling with the problem at hand, they generate solution ideas that most likely are incomplete or erroneous (e.g., Kapur & Bielaczyc, 2012). However, these erroneous or incomplete student solutions can form the basis for acquiring valid knowledge during subsequent instruction. An instructional approach that resemble the described scenario is the so-called productive failure approach (Kapur, 2010, 2012): After an initial problem-solving phase, the teacher introduces and explains the correct solution and the underlying concept during the subsequent instruction phase. Multiple studies have shown the effectiveness of productive failure with respect to the acquisition of conceptual knowledge in comparison to an instructional design with reverse order (i.e., instruction followed by problem-solving) (e.g., Kapur, 2010, 2012, 2014; Kapur & Bielaczyc, 2012; Loibl & Rummel, 2014; for a review see: Loibl, Roll, & Rummel, 2017). The beneficial effect of the productive failure approach is attributed to the activation of prior knowledge and intuitive ideas in the initial problem-solving phase (e.g., Kapur & Bielaczyc, 2012). Prior knowledge activation, in turn, should help students to integrate new information received during the subsequent instruction phase in their prior knowledge structure (e.g., Sweller, 1988).

However, research showed that the problem-solving activity remains ineffective if followed by an instruction that focusses only on correct solutions (Loibl & Rummel, 2014). In this study, students in the productive failure condition with "problem solving prior to instruction" outperformed their counterparts in the reverse condition "instruction prior to problem solving" with respect to conceptual knowledge only when the instruction built on erroneous student solutions by comparing the erroneous student solutions to the correct solution. Thus, it seems that prior knowledge activation by itself does not fully explain the effectivity of productive failure. In contrast, the finding of this study stresses the importance of building on erroneous solution attempts during the instruction phase.

This finding fits the literature on conceptual change and learning from errors: Comparing erroneous to correct examples focuses students' attention on the components that differ (e.g., Durkin & Rittle-Johnson, 2012). Thus, comparing erroneous solution attempts to the correct solution focuses students' attention on aspects that they still have to learn. In other words, students' knowledge gaps are specified (Loibl & Rummel, 2014). As learning takes place once students realize that they reached an impasse (VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003), students need to become aware of their specific knowledge gaps in order to revise their mental model (e.g., Chi, 2000; Vosniadou & Verschaffel, 2004).

In previous studies on productive failure, the teacher (or experimenter) led the comparison between erroneous solution attempts and the correct solution. However, what influences the learning outcomes are the cognitive activities of the learners. Therefore, a relevant open question (that we target in our study) is whether prompting students to engage in such comparisons by themselves can foster learning. Prompts elicit specific learning processes (Renkl 2005), often by fostering self-explanations that learners would not provide by themselves (Pressley, Wood, Woloshyn, Martin, King, & Menke, 1992). Chi and colleagues even found benefits of prompted self-explanations in comparison to spontaneous self-explanations (Chi, de Leeuw, Chiu, & Lavancher, 1994). Against this background, it is not surprising that prompts have been shown as effective means to foster many kinds of learning activities (e.g., Berthold, Nückles, & Renkl, 2007; Devolder, van Braak, & Tondeur, 2012; Schworm & Renkl, 2007). It therefore seems to be a reasonable approach to include prompts during the instruction phase of productive failure in order to support students' learning processes.

Also, it seems reasonable to argue that triggering these comparison processes is most important for topics with epistemological barriers which require a conceptual change (e.g., Vosniadou & Verschaffel, 2004). One well-studied topic with epistemological barriers is fractions (Prediger, 2008): Knowledge on natural numbers often cannot be applied to fractions, which leads to typical obstacles in understanding relevant aspects of the fraction concept. For instance, typical errors when comparing fractions arise when students focus only on the numerator or the denominator and, thus, produce erroneous graphical and symbolical representations (Eichelmann, Narciss, Schnaubert, & Melis, 2012; McNamara & Shaughnessy, 2011; Stafylidou & Vosniadou, 2004).

As indicated earlier, productive failure has been shown effective for fostering the acquisition of conceptual knowledge. Conceptual knowledge is defined as understanding about underlying principles and structures of a domain (cf. Rittle-Johnson & Alibali, 1999). In other words, conceptual knowledge includes an understanding of the underlying concepts of a solution and it allows reasoning about why and how a procedure works. In addition, there is reason to believe, that comparing incorrect and correct solution attempts during the learning phase fosters the acquisition of so-called "negative knowledge". Negative knowledge refers to knowing what is not part of a concept and what procedure does not work and why (Oser, Hascher, & Spychiger, 1999). Thus, negative knowledge draws a line between correct and incorrect solution attempts and prevents students from making mistakes (again) that are covered by their negative knowledge. Heemsoth and Heinze (2014) found that reflecting on incorrect examples indeed supported students' negative knowledge more than reflecting on correct examples. However, in their study this increase of negative knowledge came at the cost of lower conceptual knowledge (in comparison to students that reflected on correct examples) for students with low prior knowledge. Therefore, potentially differential effects on conceptual knowledge and negative knowledge need to be considered.

Research question

Against this background, our main research question is: Does prompting comparisons between incorrect and correct solution attempts after a problem-solving activity foster the acquisition of conceptual knowledge and negative knowledge about fractions? For conceptual knowledge and negative knowledge, we hypothesize that students who are prompted to compare incorrect and correct solution attempts will outperform their counterparts who are not prompted and/or do not work with incorrect solution attempts at all (hypothesis 1).

With regard to the learning process, we investigate, whether students engage in elaborations on errors and comparisons with and without being prompted to do so. We hypothesize that most prompted students will engage in theses intended processes, while only few students who are not prompted will do so (hypothesis 2).

Methods

Participants

Participants were 200 fifth-graders from nine classes in Germany. Individual students were randomly assigned to three conditions.

Learning material

The topic fractions had not been covered in class prior to the study. The learning unit of our study covered comparing fractions with graphical and numerical representations. The material relied on the experiences and findings from the KOSIMA project (KOntexte für SInnstiftendes MAthematiklernen [contexts for meaningful mathematics learning]). The KOSIMA project developed and implemented teaching units for mathematics, including fractions. In our study, we built on core elements of the fractions material from Prediger, Barzel, Hußmann, and Leuders (2013). For more details on the findings on learning fractions with the selected material, see Prediger, Glade, and Schmidt (2011).

In the *problem-solving phase*, all students were asked to decide which team wins a scoring contest where each player attempts to score a goal once: a team of 5 girls who scored a total of 3 goals or a team of 10 boys who scored a total of 5 goals. It was clarified that each team member only had one attempt as displayed in Figure 1. The first relevant knowledge component is to understand that the absolute number of goals does not lend itself to a fair comparison as the number of children (i.e., the number of attempts) differs between the groups. Building upon this understanding, the number of goals has to be set in proportion to the number of group members: 3/5 and 5/10. Finally, these fractions with differing denominators need to be compared. This can be done on a graphical level with fraction bars or on a symbolic level by expanding 3/5 with 2.

Fairly comparing number of goals

Two teams compete in a scoring contest where each player attempts to score a goal. Each player has only one attempt. The table provides the results of the two teams with each square representing one goal.

	Goals
Pia`s team (5 girls)	☐☐☐
Ole`s team (10 boys)	☐☐☐☐☐

Compare the results of the two groups. Who wins – girls or boys? Draw a graphical representation (e.g., boxes or bars) to explain your decision. Label your representation with fractions.

Figure 1. Translated task in the problem-solving phase.

In the *instruction phase*, all students received two canonical solutions marked as correct. The students in the error condition and in the prompt condition additionally received two erroneous solution attempts (focus on number of goals only; comparing number of goals and number of group members separately) marked as incorrect. Figure 2 provides an example of an erroneous solution attempt.

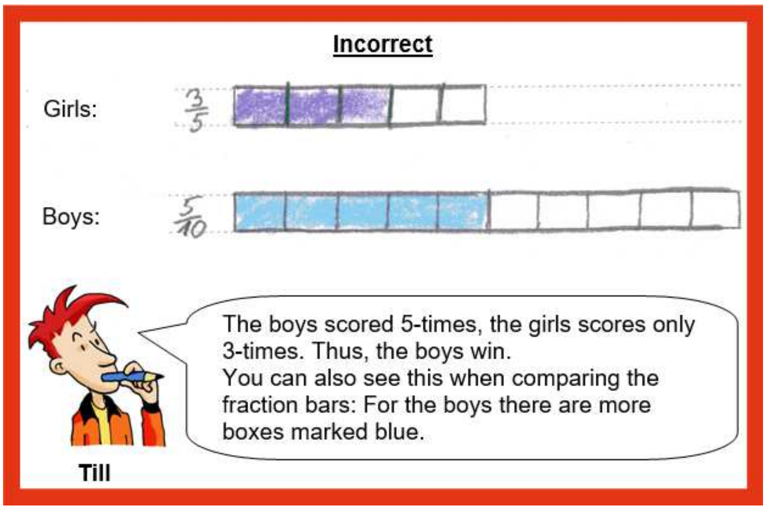

Figure 2. Translated example of an erroneous solution attempt provided in the instruction phase.

All students received a written sheet that specified their tasks. This sheet included different prompts for the different conditions: Students in the control condition and in the error condition received the following general prompt: “Have a look at the solution ideas. Was the comparison fair? Who wins?” The sheet did not specify whether the students should elaborate on the solution attempts one by one or overall. Students in the prompt condition received more specific prompts to compare two solution attempts (a correct and an incorrect attempt) at a time, to identify the error in the erroneous solution attempt, and to explain what is better in the correct solution (e.g., "Compare the solution ideas of Till and Ole. What did Till do wrong? What did Ole do better than Till?"). Finally, all students were asked what they need to consider in general in order to compare fairly when the number of group members differs (e.g, when 3 kids compete against 9 kids).

Measurements

Prior to the study, we administered a prior knowledge test, a test regarding students' mathematical operations sense, and demographical data. At the end of the study, we administered a posttest testing for conceptual knowledge and negative knowledge.

The test for *prior knowledge* covered items related to the learning unit (isomorph to the conceptual knowledge posttest items, see below). The prior knowledge test did not include any items targeting negative knowledge. Including negative knowledge items (cf. posttest items below) in the pretest would have altered the exploration phase by providing solution ideas and would have interfered with our manipulation by providing erroneous solution attempts to all students.

The *mathematical operations sense* was tested with items adopted from Lernstand 5, a central comparative test (Schulz, Leuders, & Rangel, 2017). The test included 10 items such as "Pia's dad is 48 years old. He is 6-times as old as Pia. How old is Pia? Please write down your calculation (not the result)."

Finally, students filled in a short questionnaire on *demographical data* asking for gender, age, first language, and mathematics score of the last school report.

The *posttest* included items on conceptual knowledge and items on negative knowledge. The items on conceptual knowledge introduced problems similar to the one in the problem-solving activity and asked students to explain their reasoning. The negative knowledge items presented student solutions to similar problems and asked students to identify whether these solution attempts are correct or incorrect and to explain their reasoning. The maximum was 7 points for conceptual knowledge (6 points for correctly identifying the biggest fraction and reasoning in each case plus 1 point for general reasoning) and 12 points for negative knowledge (for identifying six solution attempts as correct or incorrect and reasoning in each case).

Regarding our *process measure*, we analyzed students' answers to the elaboration task in the instruction phase. We coded a) whether students explicitly referred to at least one of the erroneous solution attempt and b) whether they explicitly identified the error. Students from the control condition were excluded from this analysis because they did not receive erroneous solution attempts during the instruction phase and, thus, were not able to refer to them in this phase.

Procedure

In order to keep the instruction equal across conditions, all classes were taught by trained teachers, who did not know our hypotheses, but were aware of the differences in the learning material. The study started with an introductory session for all conditions to guarantee a basic understanding of fractions (e.g., division of 4 pizzas for 6 children without remainder). This session was followed by a prior knowledge test.

During the main session of the study, all students first worked for 25 minutes individually on a problem on comparing fractions with unequal denominators (problem-solving phase). The previous introductory session covered neither comparing fractions nor expanding fractions (in order to obtain a common denominator). Thus, the target concept was unknown to the students. The problem-solving activity was identical across conditions. Students handed in their solution attempts at the end of this phase.

Afterwards students worked for 20 minutes on elaboration tasks which introduced canonical solutions (instruction phase). This phase differed according to condition as displayed in Table 1.

Table 1: Differences between conditions during the instruction phases.

Condition	Correct solutions	Erroneous solution attempts	Comparison prompt
Control condition	Yes	No	No
Error condition	Yes	Yes	No
Prompt condition	Yes	Yes	Yes

In the *control condition*, student received two canonical solutions to the problem. They were asked to elaborate on the solution attempts and to explain how to solve such a problem in general. In the *error condition*, students additionally received two typical erroneous solution attempts marked as such. The elaboration tasks were the same as in the control condition. Thus, in the error condition students were able to draw comparisons between correct and incorrect solution attempts on their own, but they were not explicitly prompted to do so. In the *prompt condition*, students received the correct and erroneous solution attempts and were additional prompted to compare two solution attempts at a time before answering the elaboration task. Students of all conditions were asked to write down their answers to the elaboration task. There answers (as well as the correct and erroneous solution attempts) were collected at the end of the instruction phase. Students of all conditions worked individually in the same classroom. Time on task was held constant across condition.

Afterwards all students completed a posttest targeting conceptual knowledge and negative knowledge. The posttest took 25 minutes. All students finished the test in time.

Results

Prior knowledge

The conditions did not differ in their prior knowledge on comparing fractions [$F(2,194) = 0.234$, $p = .79$], regarding their mathematical operation sense [$F(2,193) = 0.153$, $p = .86$], nor regarding their general math score from the last school report [$F(2,151) = 0.330$, $p = .72$]. Table 2 displays an overview of these control variables.

Table 2: Means (SD) of the control variables.

Condition	Prior knowledge on comparing fractions (max. 7 points)	Mathematical operation sense (max. 10 points)	General math score (1 to 6, 1 is the best score)
Control condition	1.58 (2.01)	5.55 (2.14)	2.15 (1.17)
Error condition	1.36 (1.67)	5.75 (2.12)	2.18 (1.19)
Prompt condition	1.49 (1.93)	5.66 (2.00)	2.31 (1.05)

Learning outcomes

About 20% of the posttest data was coded by a second independent rater. Interrater-reliability (ICC) was high for both scales, conceptual knowledge ($ICC_{2,1} = .993$) and negative knowledge ($ICC_{2,1} = .954$). Only the 196 students with complete datasets were included in the following analyses. Table 3 presents the mean test scores for the three conditions.

Table 3: Means (SD) of the posttest scores.

Condition	N	Conceptual knowledge (max. 7 points)	Negative knowledge (max. 12 points)
Control condition	64	3.03 (2.56)	4.27 (2.72)
Error condition	64	3.33 (2.22)	4.75 (2.67)
Prompt condition	68	3.81 (2.47)	5.47 (2.90)

Prior knowledge on comparing fractions (conceptual knowledge: $r = .491$, $p = .000$, negative knowledge: $r = .220$, $p = .002$) and the mathematical operation sense (conceptual knowledge: $r = .445$, $p = .000$, negative knowledge: $r = .291$, $p = .000$) significantly correlated with the posttest scores and were therefor include as covariates. ANCOVAs with prior knowledge and mathematical operation sense as covariates revealed (marginal) significant effects for conceptual knowledge [$F(2,191) = 2.70$, $p = .07$, $\eta_p^2 = .03$] and for negative knowledge [$F(2,191) = 3.46$, $p = .03$, $\eta_p^2 = .04$]. Posthoc comparisons (LSD) revealed significant differences between the prompt condition and the control condition, both for conceptual knowledge ($p = .02$) and for negative knowledge ($p = .01$), favoring the prompt condition. Differences between the error condition and the control condition (conceptual knowledge: $p = .34$, negative knowledge: $p = .31$) and between the error condition and the prompt condition (conceptual knowledge: $p = .19$, negative knowledge: $p = .12$) were not significant. Taken together, our results partially support hypothesis 1: The prompt condition significantly outperformed the control condition on both scales, but there was no significant effect compared to the error condition.

 © ISLS

Process data

Our process analyses yielded insights on whether students in the error condition (without being prompted) and students in the prompted condition did attend to the erroneous solution attempts. From the 65 students in the error condition only 6 referred to an erroneous solution attempt and from these 6 only 2 identified an error. In contrast, in the prompt condition 67 of 68 students referred to the erroneous solution attempts and 56 of them identified at least one error. Thus, our results descriptively support hypothesis 2.

Discussion

The research presented here focuses on the instructional approach productive failure (cf. Kapur & Bielaczyc, 2012; Loibl et al., 2016). Productive failure combines the generation of divergent solution attempts in a problem-solving phase with a subsequent instruction phase. Our study focused on the processes in the instruction phase. More precisely, we investigated whether elaborating on erroneous solution attempts and comparing them to correct solutions after problem solving fosters learning fractions. To answer this question, we compared three conditions: Students of all experimental conditions first engaged in an identical exploration activity. In the subsequent instruction phase they worked on an elaboration task that introduced either a) only correct solutions (control condition), or b) correct solutions and typical erroneous solution attempts without prompts (error condition) or c) correct solutions and typical erroneous solution attempts with prompts to compare these solution attempts (prompt condition).

Our results indicate that including erroneous solution attempts is beneficial for learning (both conceptual knowledge and negative knowledge), if students are explicitly prompted to compare them to correct solutions (prompt condition): At posttest, students in the prompt condition significantly outperformed students in the control condition who studied correct solutions only.

In contrast to our results, Heemsoth and Heinze (2014) found beneficial effects of elaborating on errors on negative knowledge and not on conceptual knowledge. These divergent findings might be explained by the fact that the study by Heemsoth and Heinze compared two slightly different conditions: elaboration on correct examples only (cf. our control condition) versus elaboration on incorrect examples only. None of the conditions engaged in comparisons between correct and incorrect examples (cf. our prompt condition). Thus, our results indicate that including erroneous solution attempts is only beneficial for the acquisition of conceptual knowledge, if students are explicitly prompted to compare them to correct solutions.

In our study, students who received the erroneous solutions without comparison prompts (error condition) did not differ significantly from the other conditions. To some extent, this result supports the notion that confrontation with errors alone does not unfold the full potential of learning from errors (cf. Große & Renkl, 2007; Heemsoth & Heinze, 2014). As the error condition and the control condition did not differ significantly on any tests, it is of interest whether students in the error condition actually attended to the erroneous solution attempts. While the implicit cognitive activities naturally remain unrevealed, our process data shows that almost no student in the error condition explicitly elaborated on the errors. In other words, the students in the error condition did not take advantage of the opportunity to elaborate on errors (cf. hypothesis 2) and their learning process therefore resembles the process of students in the control condition (at least on the surface level). This finding underlines the importance of differentiating between the intended learning process and the process that actually takes place (cf. Dillenbourg, Baker, Blaye, & O'Malley, 1996). In contrast to the error condition, almost all of the prompted students explicitly referred to the erroneous solutions and identified the errors while working on the elaboration task. Thus, in line with our hypothesis 2 our prompts triggered the intended processes that, in turn, seem to be beneficial for learning. However, the link between the process and learning is rather speculative, because the prompt condition did not significantly outperform the error condition at posttest. Future research should investigate in more details the different processes and their relation to learning.

While attempting to keep materials as parallel as possible, we have to acknowledge that the number of tasks and prompts differed between conditions. Despite this difference, time on task was held constant across condition. This was possible as the conditions with fewer prompts (i.e., the error condition and the control condition) received broad prompts, while the condition with more prompts (i.e., the prompt condition) received very specific prompts. Our hypotheses focus the type of prompts. However, the number of prompts also may have affected the learning outcomes. Future research may detangle the effects of number of prompts and type of prompts.

Inspired by the in vivo research paradigm advocated of the Pittsburgh Science of Learning Center (Koedinger, Corbett, & Perfetti, 2012), we conducted our study in schools during regular mathematics lessons. Nevertheless, we kept the internal validity high by highly standardizing our procedure. However, this highly controlled study design comes at a cost: Except of the experimental manipulation, we kept the instruction phase equal for all students. That is, all students received the very same correct and incorrect solution attempts, independent from what they themselves produced during the problem-solving phase. Potentially, a flexible instruction that reacts to individual solution attempts of students in a more specific and adaptive way could be even more effective. We aim to test this hypothesis in a future study by using a computer-based adaptive environment.

References

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction, 17*(5), 564-577.

Chi, M. T. H. (2000). Self-explaining expository texts: the dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in Instructional Psychology* (pp. 161-238). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

Devolder, A., van Braak, J., & Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: systematic review of effects of scaffolding in the domain of science education. *Journal of Computer Assisted Learning, 28*(6), 557-573.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In H. Spada & P. Reimann (Eds), *Learning in Humans and Machine: Towards an interdisciplinary learning science* (pp. 189-211). Oxford: Elsevier.

Durkin, K., & Rittle-Johnson, B. (2012). The effectiveness of using incorrect examples to support learning about decimal magnitude. *Learning and Instruction, 22*(3), 206-214.

Eichelmann, A., Narciss, S., Schnaubert, L., & Melis, E. (2012). Typische Fehler bei der Addition und Subtraktion von Brüchen – Ein Review zu empirischen Fehleranalysen [Typical error when addiding and subtracting fractions – A review on empirical analyses on errors]. *Journal für Mathematik-Didaktik, 33*(1), 29-57.

Heemsoth, T., & Heinze, A. (2014). The impact of incorrect examples on learning fractions: A field experiment with 6th grade students. *Instructional Science, 42*(4), 639-657.

Große, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17*(6), 612-634.

Kapur, M. (2010). A further study of productive failure in mathematical problem solving: Unpacking the design components. *Instructional Science, 39(4)*, 561-579.

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science, 40*(4), 651-672.

Kapur, M . (2014). Productive failure in learning math. *Cognitive Science, 38*(5), 1008-1022.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences, 21*(1), 45–83.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning. *Cognitive Science, 36*(5), 757-798.

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a Theory of When and How Problem Solving Followed by Instruction Supports Learning. *Educational Psychology Review, 29*(4), 693-715.

Loibl, K., & Rummel, N. (2014). Knowing what you don't know makes failure productive. *Learning and Instruction, 34,* 74-85.

McNamara, J., & Shaughnessy, M. M. (2011). *Student errors: what can they tell us about what students do understand?* Available at http://www.mathsolutions.com/documents/studenterrors_jm_ms_article.pdf (retrieved June 2017).

Oser, F., Hascher, T., & Spychiger, M. (1999). Lernen aus Fehlern. Zur Psychologie des „negativen“ Wissens [Learning from errors. About the psychology of negative knowledge]. In W. Althof (Ed.), *Fehlerwelten. Vom Fehlermachen und Lernen aus Fehlern [Worlds of Errors. On Making errors and learning from errors]* (pp. 11-41). Opladen: Leske + Budrich.

Prediger, S. (2008). The relevance of didactical categories for analysing obstacles in conceptual change - Revisiting the case of multiplication of fractions. *Learning and Instruction, 18*(1), 3-17.

Prediger, S., Glade, M., & Schmidt U. (2011). Wozu rechnen wir mit Anteilen? Herausforderung der Sinnstiftung am schwierigen Beispiel der Bruchoperationen [Why calculating with fractions? Challenges of creating meaning using the difficult example of fractions]. *Praxis der Mathematik in der Schule, 53*(37), 28-35.

Prediger, S., Barzel, B., Hußmann, S., & Leuders, T. (2013). *mathewerkstatt 6*. Berlin: Cornelsen

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27*, 91-109.

Renkl, A. (2005). The worked-out example principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 229-247). Cambridge, UK: Cambridge University Press.

Rittle-Johnson, B., & Alibali, M.W. (1999). Conceptual and Procedural Knowledge of Mathematics: Does One Lead to the Other? *Journal of Educational Psychology, 91,* 175-189.

Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology, 99*(2), 285-296.

Schulz, A., Leuders, T., & Rangel, U. (2017). Empirie- und modellgestützte Diagnostik von arithmetischen Basiskompetenzen als Grundlage für Förderentscheidungen zu Beginn von Klasse 5 [Empirical and model-based diagnosis of basic arithmetical competences as basis for decision on support at the beginning of grade 5]. In A. Fritz-Stratmann & S. Schmidt (Eds.), *3. Handbuch Rechenschwäche [Handbook Dyscalculia]* (pp. 396-417). Weinheim: Beltz.

Stafylidou, S., & Vosniadou, S. (2004). The development of students' understanding of the numerical value of fractions. *Learning and Instruction, 14*, 503-518.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). *Why do only some events cause learning during human tutoring? Cognition and Instruction, 21*(3), 209-249.

Vosniadou, S., & Verschaffel, L. (2004). Extending the conceptual change approach to mathematics learning and teaching. *Learning and Instruction, 14*(5), 445-451.

Acknowledgments

This research was supported by the Deutsche Forschungsgemeinschaft, DFG (LO 2196/1-1). We would like to thank the participating schools and students. Thanks to our student assistants for their help in collecting and coding the data.