Received: 8 September 2019 | Revised: 23 February 2020 | Accepted: 1 March 2020
DOI: 10.1002/acp.3651

SPECIAL ISSUE ARTICLE

WILEY

Do examples of failure effectively prepare students for learning from subsequent instruction?

Christian Hartmann[1] | **Tamara van Gog**[2] | **Nikol Rummel**[1]

[1]Institute of Educational Research, Ruhr University Bochum, Bochum, Germany
[2]Department of Education, Utrecht University, Utrecht, The Netherlands

Correspondence
Christian Hartmann, Institute of Educational Research, Ruhr University Bochum, Universitätsstraße 150, 44780 Bochum, Germany.
Email: christian.hartmann@rub.de

Summary

Studies on the productive failure (PF) approach have demonstrated that attempting to solve a problem prepares students more effectively for later instruction compared to observing failed problem-solving attempts prior to instruction. However, the examples of failure used in these studies did not display the problem-solving-and-failing *process*, which may have limited the preparatory effects. In this quasi-experiment, we investigated whether observing someone else engaging in problem solving can prepare students for instruction, and whether examples that show the problem-solving-and -failing *process* are more effective than those that only show the outcome of this process. We also explored whether the perceived model–observer similarity had an impact on the effectiveness of observing examples of failure. The results showed that observing examples effectively prepares students for learning from instruction. However, observing the model's problem-solving-and-failing process did not prepare students more effectively than merely looking at the outcome. Studying examples were more effective if model–observer similarity was high.

KEYWORDS

conceptual knowledge acquisition, example-based learning, mathematics, observational learning, productive failure, vicarious failure

1 | INTRODUCTION

Research on the productive failure approach (PF; Kapur, 2012) has shown that attempting and "failing" to solve a problem prior to instruction is more effective for improving students' conceptual knowledge acquisition (i.e., their understanding of the deep features of a concept) than a direct-instruction (DI) approach in which students first receive instruction and then apply the learned problem-solving strategies (Kapur, 2012). Moreover, this benefit seems to occur without hampering the acquisition of procedural knowledge, that is, the knowledge regarding how to apply, for instance, a mathematical formula (for an overview, see Loibl, Roll, & Rummel, 2017). This preparatory effect of problem solving has been especially demonstrated for mathematics and science education (Darabi, Arrington, & Sayilir, 2018).

Even if students who are engaged in problem-solving attempts prior to instruction do not yet know the concept required to solve a thematically related problem, attempting to solve the problem seems to make them more receptive to the instruction (Loibl et al., 2017). Research on PF explains this preparatory effect by means of *prior knowledge activation* and *awareness of knowledge gaps* evoked through attempting and failing to solve a problem prior to instruction (Loibl et al., 2017). For instance, if students first activate prior knowledge through problem-solving attempts, this should help them to cognitively process the information about the targeted concept provided in the subsequent instruction, that is, to organize and integrate the new information with prior knowledge. In addition, it is assumed that the PF approach enables students to develop an awareness of their knowledge gaps (Loibl & Rummel, 2014a). If students initially fail and become aware of their knowledge gaps, it might be easier for them to identify, organize, and ultimately integrate the missing knowledge which is subsequently imparted during instruction. Moreover, reaching an impasse and becoming aware of knowledge gaps is assumed to make students more curious about how to solve the problem, consequently making them more motivated to learn from the later instruction.

Nevertheless, despite studies showing that the PF approach is more effective than DI, and some indications that the aforementioned mechanisms might explain this effect (Loibl et al., 2017), it remains unclear which preparatory activities during problem solving prior to instruction indeed promote conceptual knowledge acquisition. It seems reasonable to assume, however, that there may be alternative ways to prepare students for instruction, for instance by having them *observe someone else* engaging in problem-solving attempts, that is, studying examples of failure. To investigate whether students need to generate their own problem-solving attempts, Kapur (2014a, 2014b) compared a PF condition to a "vicarious failure" (VF) condition, in which students were prompted to assess failed solution attempts generated by other students before receiving instruction, as well as to a DI condition. The results showed that students in the PF condition outperformed those in the VF and DI conditions in a conceptual knowledge post-test. Kapur (2014b) concluded that generating their own solution attempts prepares students more effectively due to a better activation of prior knowledge, as the vicarious experience of failure was less effective than PF. However, students in the VF condition also outperformed students in the DI condition. Accordingly, observing another student's outcomes of problem-solving attempts seems to trigger at least some preparatory mechanisms similar to PF, although not as effectively as PF.

A possible explanation for why VF was better than the DI control condition, but not as effective as PF in Kapur's (2014a, 2014b) studies, might lie in the fact that only the outcome of the attempt, and not the entire process was shown. Accordingly, students looked at final solution attempts (i.e., problem-solving *outcome*), but without seeing and hearing the actions and thoughts of the model as they unfolded throughout the problem-solving *process*, that is, the model's intentions or conclusions regarding the various steps in the process. Furthermore, as the examples used by Kapur (2014a, 2014b) were didactically prepared to make them more comprehensive, they appeared less like authentic student examples. This lack of authenticity could have decreased the observers' personal involvement (cf. Couzijn, 1999) and their identification with the model, thus preventing them from adopting the model's failure as their own and diminishing the success of VF. In light of the literature on example-based and observational learning, it seems reasonable to expect that the effectiveness of observing examples of another student who is attempting and failing to solve a problem would depend on, firstly, which parts of the problem-solving process the observer has access to, and secondly, the perceived model–observer similarity, that is, the extent to which the observer identifies with the model whose solution attempts he or she is studying. If examples from failing students indeed provide more insight into the model's problem-solving process (i.e., by using video modeling examples) and enable observers to identify with the PF model more strongly, this might lead to a more effective activation of the observer's own prior knowledge and a better awareness of his/her own knowledge gaps. Accordingly, studying examples of failure should be more effective than in Kapur's (2014a, 2014b) study. We will elaborate on these topics in the following two sections.

The present quasi-experimental study compares VF conditions in which students either observed the *process* of another student engaged in problem-solving attempts or only looked at the final outcome of the model's problem-solving to a condition in which students were instructed first (DI), that is, the standard control condition as used in classical PF studies (e.g., Kapur, 2014a, 2014b; Loibl & Rummel, 2014a, 2014b). We additionally explored the potential impact of model-observer similarity. Besides contributing to research on the PF paradigm, the present study also contributes to research on example-based and observational learning. Research on example-based learning demonstrates the beneficial effects of studying worked examples (i.e., of the canonical solution or parts thereof) prior to instruction (for review, see Van Gog, Rummel, & Renkl, 2019), and of studying erroneous examples subsequent to instruction (e.g., Tsovaltzi, McLaren, Melis, & Meyer, 2012). However, so far, it is unclear whether studying *examples of failure prior to receiving instruction* can also be conducive to learning, and if so, under which conditions.

2 | VIDEO MODELING EXAMPLES: MAKING FAILURE OBSERVABLE

As mentioned above, Kapur (2014a, 2014b) converted original student solutions into "well-designed" static worked examples. The VF students in his study looked at "typical" problem-solving outcomes, augmented with a brief description of what a solution was about, and then had to evaluate whether the provided examples displayed good or weak problem-solving attempts. Therefore, VF students did not observe the "natural" behavior as performed by another student during the problem-solving process. This natural model behavior, which can be made visible by using video modeling examples (for a discussion of worked examples and modeling examples, see Van Gog & Rummel, 2010; Van Gog et al., 2019) or auditory think-aloud recordings, might, however, have included parts of the problem-solving process which make the PF approach effective for conceptual knowledge acquisition.

Recordings of the problem-solving process of the model, for instance, might reveal detailed information about how the model explored and interpreted the problem-solving task. Importantly, think-aloud recordings of the model's thoughts would make it more salient to the observer why the model came up with a certain solution attempt, accompanied by further descriptions. In other words, the observer would have access to how the model used prior knowledge to come up with a solution attempt, and this would help the observer to activate his/her own prior knowledge as well. Furthermore, experiencing incompetence or failure is assumed to be an essential preparatory effect of engagement in problem solving prior to instruction. For instance, Loibl and Rummel (2015) found that students who received instruction prior to problem solving overestimated their competence after instruction. By contrast, students who engaged in problem-solving attempts prior to instruction perceived a gain in competence from the preparation activity to instruction, acquired more conceptual knowledge from the instruction, and assessed their learning more accurately (Loibl & Rummel, 2015). Accordingly, existing research on the PF paradigm suggests that the experience of failure, that is, the perception of incompetence is a relevant preparatory effect. It seems reasonable to assume that observing another student engaged in problem-solving attempts would also trigger the experience of failure, however, it is unclear upon which criteria observers assess their competence and experience failure when they are not actually failing themselves.

If students attempt to solve a problem, which is completely unfamiliar to them, they will most likely have fewer objective criteria or standards on hand to accurately assess problem-solving performance. Students who engage in their own problem-solving attempts might experience failure because they have the impression that they are unable to solve the problem by means of their own prior knowledge and abilities. In other words, they might detect flaws in their solution attempts, making them aware of their knowledge gaps or incompetence (Loibl & Rummel, 2014a). It stands to reason that students who do not engage in problem solving themselves but observe problem-solving attempts may likewise develop a better awareness of their own knowledge gaps. They may identify with the model (vicarious experience of failure) and—at least when they can observe the process—they also observe how the model elaborates on flaws (or strengths) of his or her solution attempts. In contrast, students who are only provided with the finished solutions, without any further process information, may not develop such awareness.

In conclusion, the lack of information about the problem-solving process of the model in Kapur's studies (2014a, 2014b) might explain why VF was less effective compared to PF, because the effectiveness of observing someone else engaged in problem-solving attempts also depends on the extent to which the model's failure whereas engaged in problem solving is observable. By giving observers access to the model's intentions, expectations, conclusions, and critical reflections, video modeling examples might more effectively trigger observers' own prior knowledge activation and awareness of knowledge gaps. Thus, information about how the model actively explored the not-yet-known concept might be the key for preparing students more effectively for learning from instruction by means of examples of failure.

3 | MODEL–OBSERVER SIMILARITY

Besides the observability of the problem-solving process, the characteristics of the model, and especially the observer's perception of these characteristics, might also influence the effectiveness of observing another student engaged in problem-solving attempts prior to instruction. Research on observational learning has indicated that if observers identify strongly with the model, they may be more likely to see the model's failure as also applicable to themselves. According to research on modeling, and in particular on model–observer similarity (Schunk, 1987), "the greater the assumed similarity, the more persuasive are the model's success and failures" (Bandura, 1994, p. 81). If an observer feels that he/she is more competent than the model, for instance, the observer may dismiss the model's failure as irrelevant (i.e., "I would have done better") and pay less attention to the examples and subsequent instruction. Thus, regardless of how good or bad the displayed solution attempts actually are, the perceived similarities (high identification) or dissimilarities (low identification) between the observer and the model may affect the observer's experience of failure, how the observer elaborates on the model's problem solving, and accordingly the preparatory effects by means of observing.

However, model–observer similarity can occur on several dimensions (e.g., competence, age, gender; Schunk, 1987) and findings on the effects of model–observer similarity on students' learning are mixed. With regard to model competence, studies on example-based learning provide some evidence for this similarity assumption. For instance, in the domain of argumentative writing, Braaksma, Rijlaarsdam, and van den Bergh (2002) found that less competent observers learned more from less competent models, and the same applied to pairs with higher competence. This effect, however, was not replicated for more creative writing tasks (Groenendijk, Janssen, Rijlaarsdam, & van den Bergh, 2013). With respect to gender, studies on observational learning found that children are more influenced by same-sex models and that boys tend to learn more aggressive behavior by observing aggressive male models than by observing aggressive female models (Bandura, Ross, & Ross, 1963). Accordingly, an observer with the same gender of the model may be more likely to adopt the failure or success of the model for him/herself. Hoogerheide, Loyens, and Van Gog (2016) and Hoogerheide, Loyens, Jadi, Vrins, and Van Gog (2017), however, did not find evidence that the effectiveness of learning from video modeling examples depends on the model's and observer's gender when the content of the examples is otherwise completely identical, although students indeed perceived higher model–observer similarity with same-gender models (Hoogerheide et al., 2017).

In conclusion, although studies have yielded mixed findings regarding the impact of perceived model–observer similarity on students' learning outcomes, it should be noted that in the context of the PF approach, learning is assumed to take place not during the initial problem-solving phase, but rather during the subsequent instruction (Loibl et al., 2017). If experiencing failure by means of failed problem-solving attempts explains preparatory effects of PF, it might be argued that students who observe failed problem-solving attempts also need to experience failure in order to be effectively prepared for the later instruction. If observers perceive high model–observer similarity by relying on model characteristics, such as competence beliefs or gender (Schunk, 1987), they might be more likely to apply the model's failure to themselves. To the best of our knowledge, there is no previous research on the PF paradigm which explored potential effects of model–observer similarities. However, such effects might partially explain why the vicarious experience of failure in Kapur's studies (2014a, 2014b) was less effective for conceptual knowledge acquisition compared to engaging in problem solving oneself.

4 | PRESENT STUDY AND HYPOTHESES

The present quasi-experimental study investigates under which conditions observing examples of failing students (VF) effectively prepares students for learning from subsequent instruction.[1] We compared VF conditions in which students either observed the *process* of another student engaged in problem-solving attempts (VF-process) or only looked at the final *outcome* of the model's problem-solving (VF-outcome). To explore the impact of model–observer similarity, we implemented two further sub-groups in both VF conditions: *same-gender* and *cross-gender*. In the same-gender condition, students studied examples from a model with the same gender (high model–observer similarity), whereas in the cross-gender condition, they studied examples of a model with the opposite gender (low model-observer similarity). We used the gender to manipulate model–observer similarity because it was the most salient model characteristic for students in the two VF conditions. The model's gender was made salient by frequently mentioning the fake name of the model while introducing the task for VF (female: Anna; male: Tim; see also, Data S1, Appendix C). Students in the VF-process condition were also able to identify the gender by the model's voice. As used in existing studies on the PF approach, we also implemented a DI condition to replicate Kapur's (2014a, 2014b) finding that by observing examples of failure (VF) effectively prepares students to acquire conceptual knowledge from the later instruction. Importantly, existing studies on PF reported beneficial effects on conceptual knowledge acquisition, while procedural knowledge acquisition was not affected (e.g., Loibl & Rummel, 2014a). We therefore only expected effects on conceptual knowledge acquisition, and not on procedural knowledge acquisition. We formulated the following hypotheses:

Hypothesis 1 *Findings by Kapur (2014a, 2014b) showed that VF outperformed the DI approach, and we expected to replicate this effect for both VF conditions. Accordingly, we hypothesized that VF-process and VF-outcome would outperform DI on the conceptual knowledge post-test.*

Hypothesis 2 *We assumed that displaying the problem-solving process to observing students would support them in activating own prior knowledge and becoming aware of own knowledge gaps. Thus, we expected that observing the problem-solving process by means of modeling-examples (VF-process) would prepare students more effectively compared to the condition in which students only looked at the outcome of the model's problem solving (VF-outcome). Note that the VF condition used by Kapur (2014a, 2014b) is comparable to the VF-outcome condition in the present study.*

Hypothesis 3 *As awareness of knowledge gaps was highlighted as an important mechanism to prepare students to benefit from the subsequent instruction (Loibl & Rummel, 2014a), we expected that mean differences between VF-process, VF-outcome and DI in the self-reported awareness of knowledge gaps would be in line with Hypotheses 1 and 2: we expected students in VF-process to become more aware of their knowledge gaps than students in VF-outcome and that students in both VF conditions would report more awareness than students in DI.*

Hypothesis 4 *Additionally, in the VF-process and VF-outcome conditions, we hypothesized that an higher identification with the model, that is, a higher model–observer similarity, would be conducive for the vicarious experience of failure. We expected that students with same-gender models would perform better on the conceptual knowledge post-test compared with students with cross-gender models. As the model's gender was salient for both VF conditions, we expected effects of perceived model–observer similarity to be equally effective in VF-process and VF-outcome conditions.*

5 | METHODS

5.1 | Participants and design

We conducted a quasi-experimental study in a secondary school (three classes) in Germany during regular mathematics lessons. The initial sample consisted of N = 71 students, but 20 students (VF-process: n = 9, VF-outcome: n = 6; DI: n = 5) did not complete all phases of our study and thus had to be excluded from the analyses. The final sample therefore comprised N = 51 students (24 female and 27 male; age: M = 16.29, SD = 0.73). The three classes were randomly assigned to DI (one class, n = 18) or VF (two classes, n = 33). Students from the two VF classes were first pooled and then randomly assigned to the VF-process (n = 15) or the VF-outcome (n = 18) condition. A power analysis indicated that our sample size is sufficient to detect a large-sized effect (f = 0.39, $1-\beta$ = .80; G-Power Analysis), which would be sufficient to detect the effects reported by Kapur (2014a) for comparisons included in our hypotheses (e.g., VF > DI: Cohen's d = 0.80).

5.2 | Materials

5.2.1 | Problem-solving task

The problem-solving task used was adopted from Kapur (2014a, 2014b), and is the task most widely used in previous PF studies (e.g., Kapur, 2014a; Loibl & Rummel, 2014a). It is a mathematical task targeting the concept of variance (i.e., mean absolute deviation) using a cover story about soccer: Learners were prompted to identify the most consistent soccer players by studying a list of the number of goals that three soccer players scored over a 10-year period. Students who attempt to solve this problem normally produce solutions that differ in how many of the following components of the canonical solution they cover (cf. Loibl & Rummel, 2014a): sum up deviations for all data points to get a precise result, take absolute or squared deviations, that is, positive values to prevent positive and negative deviations from canceling each other out, take deviations from a fixed reference point (the mean) to avoid sequence effects, and divide by the number of data points to account for sample size. Students in both VF conditions were presented with the same problem (see Data S1, Appendix B), but had access to the solution attempts generated by another student. Students in the DI condition also worked on the same mathematical problem, but after receiving instruction on the canonical solution. As students in the DI condition have already been taught about the canonical solution, they usually applied the canonical solution and graphically represented the data of the three soccer players (see Kapur, 2014a).

5.2.2 | Examples used in the VF conditions

In order to collect the material for the examples of failure in the VF conditions, 24 students from another school (henceforth also referred to as the "PF-model group") attempted to find solutions for the mathematical problem (for a selection of exemplary solutions, see Figure 1). These students generated their solutions on tablet PCs while thinking aloud, and this process was audio- and video-recorded. The examples contained differing numbers of solution attempts presented in clear handwriting. As shown in previous studies (e.g., Loibl & Rummel, 2014b), none of the presented PF-models found the canonical solution. The solution attempts of the 24 PF-models were shown to the students in the VF-process and VF-outcome conditions on tablet PCs. Thus, students in the two VF conditions did not produce their own solution attempts, but instead studied the solutions of the PF-model group. The two VF conditions differed regarding the information to which students had access. Students in the example condition with process information (VF-process) watched a video displaying the model's problem-solving steps and were able to hear the model's voice throughout the process. In contrast, students in the VF-outcome condition were only shown pictures of the final state of the model's solution attempts, without receiving the audio and video. In both VF conditions, each solution attempt was presented for the same amount of time that the model had needed to produce the attempt.

5.3 | Measures

5.3.1 | Mathematical ability and prior knowledge (pre-test)

We measured the students' mathematical ability by asking them to indicate their two most recent grades in mathematics. Therefore, mathematical ability assessed a more overarching ability to solve mathematical problems, as indicated by past academic achievements. Students' prior knowledge was measured with a pre-test equivalent to the one used by Loibl and Rummel (2014a, 2014b), with six items (Cronbach's alpha = .11) asking students to interpret and draw graphs (three points), apply their knowledge of descriptive statistics (mean and range, three points), and to draw and interpret a boxplot (four points). Within the pre-test, the students were also asked whether they knew the canonical problem solution. Note that to avoid preparatory effects by means of prior knowledge activation, the pre-test did not specifically test the concept of variance. Therefore, the pre-test gauged relevant knowledge and a broad variety of related concepts rather than specific knowledge about the targeted concept. This might explain the low internal consistency, as the six items measured different aspects of the students' prior knowledge and it cannot necessarily be assumed that these aspects are highly correlated.

5.3.2 | Quantity and quality of solution attempts

To measure the quantity of student solutions generated in the PF-model group, we counted the number of different solution attempts, irrespective of their quality. To assess the quality of solution attempts, we adopted a coding scheme from Loibl and Rummel (2014b): Each

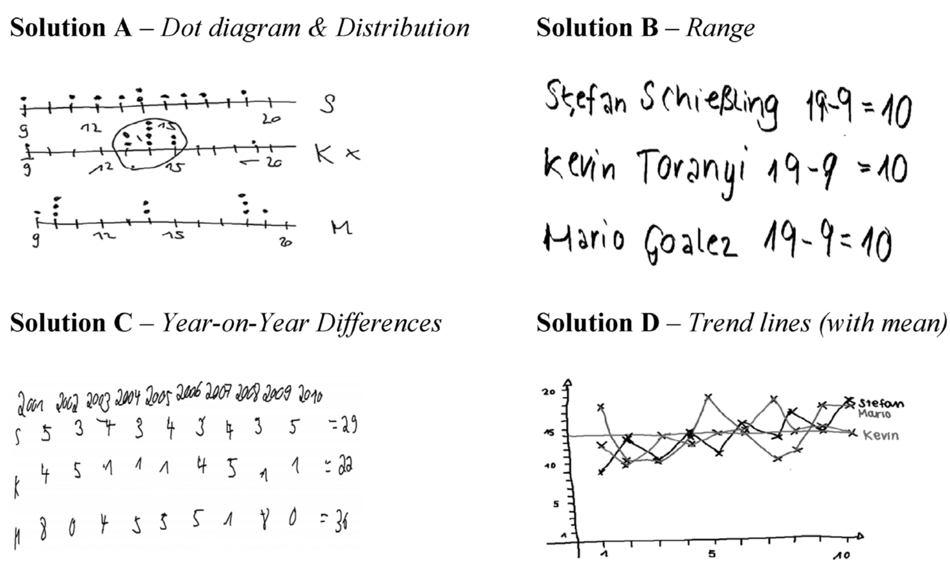

FIGURE 1 Illustration of some of the examples as used in the VF conditions

solution was further assigned to a score ranging from zero (none of the canonical solution components included) to four (all of the four canonical solution components included). To evaluate the overall quality of the solution attempts, we used the score of the one solution with the highest number of components included. A second rater coded all solution attempts generated by the 24 students of the PF-model group. The inter-rater reliability assessed by the ICC (random, absolute) was high for quantity, ICC = .94, 95% CI [.88, .98], and moderate for quality ICC = .60, 95% CI [.27, .80].

5.3.3 | Awareness of knowledge gaps

Students' self-reported awareness of knowledge gaps was measured with items adopted from Loibl and Rummel (2014a) and Glogger-Frey, Fleischer, Grüny, and Renkl (2015). Five items (Cronbach's alpha = .68) required students to rate, on a 6-point Likert scale, to what extent they perceived their own knowledge after problem solving (or observing the model or looking at the model solutions, respectively) as adequate. Some of the items asked about the produced or observed solutions more directly (e.g., "the solutions seem to be incomplete"). A high awareness of knowledge gaps was indicated if students agreed with the statement that they did not know certain things yet or had the impression of lacking knowledge. All five items can be found in the supplementary materials (Data S1, Appendix F).

5.3.4 | Conceptual and procedural knowledge (post-test)

The post-test measured conceptual and procedural knowledge and was congruent with the test used by Loibl and Rummel (2014a, 2014b). Four conceptual knowledge items (Cronbach's alpha = .44) required students to explain graphical representations of the canonical formula (two points), to sort data sets according to their distribution (two points), and to identify and explain errors of typical student solutions by relating them to the components of the canonical solutions (three points). In addition, three procedural knowledge items (Cronbach's alpha = .53) required students to procedurally apply the canonical problem-solving procedure (i.e., mean absolute deviation) to isomorphic problems (five points). A second rater coded 20% of the conceptual knowledge post-tests. Inter-rater reliability assessed by the ICC (random, absolute) was high for the entire scale, ICC = .97, 95% CI [.91, .99], as well as for each item. The post-test can be found in the supplementary materials (Data S1, Appendix E). The low internal consistency might be explained, firstly, by the low number of items, and secondly, by the fact that each item for conceptual knowledge requires students to apply their knowledge on different components of the canonical solution. Even if linking all components would indicate a high level of conceptual knowledge, a comprehensive understanding of all components would rather not be expected, especially because our results retrospectively indicate a high difficulty of the conceptual knowledge items.

5.4 | Procedure

The study consisted of four phases: Pre-test, (observing) problem solving, instruction, and post-test; the order of the (observing) problem-solving and instruction phases depended on the assigned condition (for an overview, see Figure 2). For all conditions the pre-test (duration of 20 min) took place about one week before the learning phase. There was one day between the two parts of the learning phase, that is, in the VF conditions students observed problem-solving attempts and the next day received instruction, and in the DI condition students received instruction and the next day engaged in problem solving. The post-test directly followed the learning phase, that is, in VF immediately after the instruction, in DI immediately after problem solving.

After completing the pre-test, students in the VF conditions observed another student's work on the mathematical problem for 45 min. They either observed another student's *process* of being engaged in problem-solving attempts (VF-process) or looked at the *outcomes* of a PF-model's problem-solving attempts (VF-outcome). These solution attempts came from the 24 PF-models. Students in the PF-model group had been instructed to generate as many solution attempts for the mathematical problem as possible. All of their solutions (as well as their voices while thinking aloud) were recorded on the tablet PCs, including any notes they made. Individual students from the PF-model group were then yoked to one participant from the VF-process condition and one participant from the VF-outcome condition (who would observe that PF student's solution attempts). Students in the VF-process and VF-outcome conditions were also either yoked to a student with the same (n = 16; high model–observer similarity) or opposite gender (n = 17; low model-observer similarity). The yoking procedure was designed such that (a) the assigned models had (as much as possible) similar mathematical abilities to the respective VF students and (b) the number of females and males (as well as the mathematical ability) was balanced within the same- and cross-gender subgroups. A detailed description of the yoking procedure can be found in the supplementary materials (Data S1, Appendix A).

In the introduction to the experimental phase, students in the two VF conditions were told that they were going to spend 45 min observing another student who was attempting to solve a problem. Students did not receive any support or instruction on the target concept or concerning relevant problem-solving strategies while working on the mathematical problem-solving task. For students in both VF conditions (and the PF-model group), this phase was followed by 45 min of instruction about the canonical solution from the experimenter. This instruction took place in the next mathematics lesson within one week, was the same as that used in the study by Loibl and Rummel (2014a, 2014b) and explained the canonical solution by distinguishing it from solution attempts which students typically produce.

In the DI condition, the students received this instruction first and then worked on the same mathematical problem-solving task that the VF students had observed (and the PF-model group had attempted to solve). Finally, students in all conditions worked on a 30-minute post-test requiring them to apply (procedural knowledge)

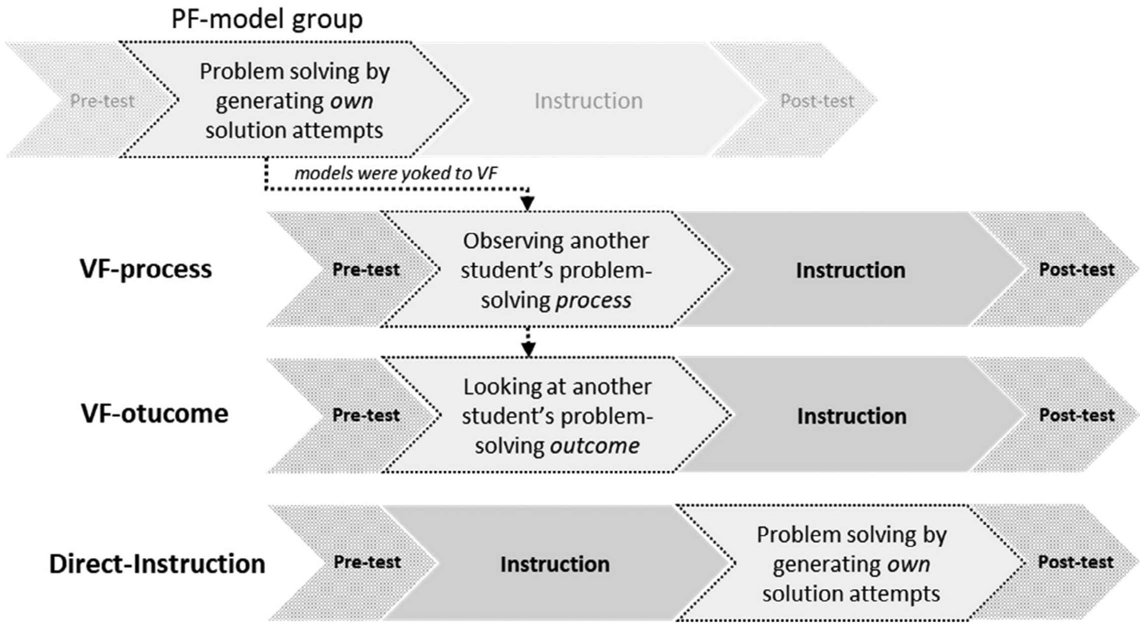

FIGURE 2 Experimental procedure

TABLE 1 Descriptive statistics

Measures		Points/scale	VF-process (n = 15) M (SD)	VF-outcome (n = 18) M (SD)	DI (n = 18) M (SD)	PF-model group (n = 24)	Total (N = 75) M (SD)
Mathematical ability[a]	Pre-test	1–6	3.13 (1.33)	3.00 (0.99)	3.11 (0.83)	3.96 (0.74)	3.36 (1.03)
Prior knowledge	Pre-test	0–10	4.27 (1.25)	3.47 (0.88)	3.81 (0.75)	4.67 (1.64)	4.09 (1.29)
Knowledge gaps	Phase 1	0–5	2.07 (0.72)	2.84 (1.04)	1.66 (0.74)	2.89 (0.88)	2.42 (1.00)
Solution quantity	Phase 1		4.27 (1.10)	4.44 (1.42)	2.44 (0.86)	4.54 (1.29)	3.96 (1.46)
Solution quality	Phase 1	0–4	2.00 (0.93)	1.78 (0.88)	3.89 (0.32)	1.83 (0.96)	2.35 (1.19)
Procedural knowledge	Post-test	0–5	3.60 (1.56)	3.67 (1.50)	3.55 (1.48)	4.10 (1.29)	3.76 (1.44)
Conceptual knowledge	Post-test	0–7	2.43 (1.62)	1.94 (1.55)	0.67 (1.03)	3.23 (1.40)	2.15 (1.69)

[a]In Germany, '1' represents the best grade while '6' represents the worst grade. To facilitate interpretation, grades were reverse-coded so that '1' represents the worst and '6' the best grade.

and explain (conceptual knowledge) the canonical solution to isomorphic problems.

6 | RESULTS

Table 1 shows the descriptive statistics (for conceptual knowledge, see also Figure 3). Further descriptive statistics as well as correlations between all variables can be found in the supplementary materials (Data S1, Appendix D).

As expected, mathematical ability correlated positively with conceptual knowledge acquisition, $r(51) = .34$, $p = .013$. Prior knowledge did not significantly correlate with the students' conceptual knowledge acquisition, $r(51) = .18$, $p = .221$. As revealed by two ANOVAs, the VF-process, VF-outcome and DI conditions did not significantly differ with regard to prior knowledge, $F(2, 48) = 2.59$, $p = .072$, $\eta_p^2 = .10$, or mathematical abilities, $F(2, 48) = 0.08$, $p = .924$, $\eta_p^2 = .003$. As the students' mathematical abilities significantly correlated with their conceptual knowledge acquisition and there were no significant differences between the conditions, mathematical ability was considered as covariate in the analyses on our hypotheses.

We defined a priori contrasts in an ANCOVA using mathematical ability as covariate to test our hypotheses on conceptual knowledge acquisition (Hypotheses 1 and 2). Overall, the ANCOVA revealed significant differences between the VF-process, VF-outcome and the DI conditions, $F(2, 47) = 8.32$, $p = .001$, $\eta_p^2 = .26$. Mathematical ability had a significant impact on conceptual knowledge acquisition, $F(1, 47) = 8.79$, $p = .005$, $\eta_p^2 = .16$.

Hypothesis 1 *The first contrast tested the hypothesis that both VF conditions (VF-process: weight of 1; VF-outcome: weight of 1) would outperform the DI condition (weight of −2) on the conceptual*

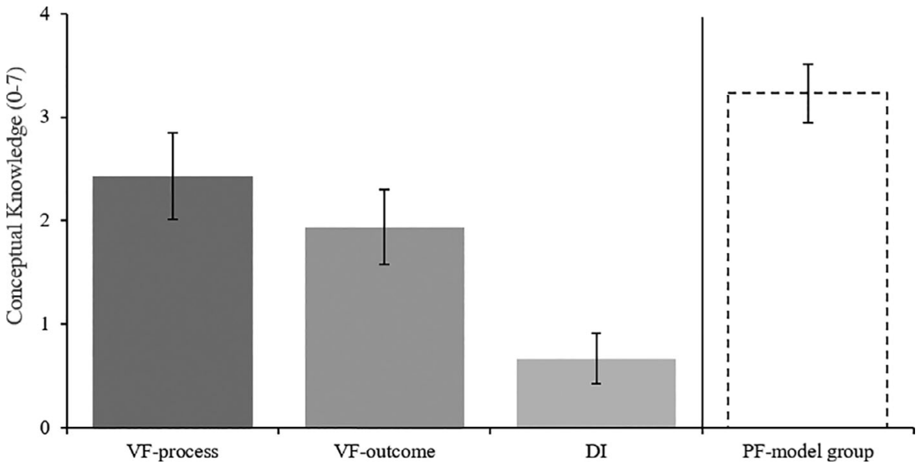

FIGURE 3 Means and standard error in conditions for conceptual knowledge acquisition

knowledge post-test. In accordance with the findings of Kapur (2014a, 2014b), students in both VF conditions performed better on the conceptual knowledge post-test than did students who first received instruction, $F(1, 47) = 16.14$, $p < .001$, $\eta_p^2 = .26$.

Hypothesis 2 *We further hypothesized that observing the problem-solving process (VF-process: weight of 1) would prepare students more effectively than the condition in which students only looked at the outcome of the model's problem-solving (VF-outcome: weight of −1). While the descriptive statistics were in the expected direction, the second contrast showed that the difference between the VF-process and VF-outcome conditions was not statistically significant,* $F(1, 47) = 0.37$, $p = .367$, $\eta_p^2 = .02$.

Hypothesis 3 *We hypothesized that the students' awareness of knowledge gaps after observing examples of failure (VF) or receiving instruction (DI) would be in accordance with Hypotheses 1 and 2. Thus, we assumed that students in the VF-process condition would become more aware of their knowledge gaps than students in the VF-outcome condition and that both VF groups would report more awareness than students who received direct instruction. To test these assumptions, an ANOVA was conducted using the same a priori contrasts as defined for Hypotheses 1 and 2. Overall, the ANOVA revealed significant differences between the conditions regarding the self-reported awareness of knowledge gaps,* $F(1, 48) = 8.94$, $p = .001$, $\eta_p^2 = .27$. *In line with our expectation, students in both VF conditions showed greater awareness after observing examples of failure than students in the DI condition,* $F(1, 48) = 10.21$, $p = .002$, $\eta_p^2 = .18$. *In contrast to our assumption, however, students in the VF-process condition reported significantly less awareness of knowledge gaps than students in the VF-outcome condition,* $F(1, 48) = 6.73$, $p = .013$, $\eta_p^2 = .12$. *Overall, the self-reported awareness of knowledge gaps did not significantly correlate with conceptual knowledge acquisition,* $r(51) = .12$, $p = .393$.

Hypothesis 4 *We hypothesized that VF-students who were assigned to PF-models with the same gender* ($n = 16$; *high model–observer similarity) would gain more conceptual knowledge compared to students who were assigned to a model with the opposite gender* ($n = 17$; *low model–observer similarity). We compared same- and cross-gender yoked students in VF with a separate ANCOVA using mathematical ability as covariate. The analysis revealed a significant effect of mathematical ability on conceptual knowledge acquisition,* $F(1, 30) = 8.72$, $p = .006$, $\eta_p^2 = .23$. *In line with our hypothesis, same-gender yoked students* ($M = 2.72$; $SD = 1.70$) *outperformed cross-gender yoked students* ($M = 1.65$; $SD = 1.30$) *in the conceptual knowledge post-test,* $F(1, 30) = 4.83$, $p = .036$, $\eta_p^2 = .14$. *Thus, students in the two VF conditions performed better on the conceptual knowledge post-test if they were more similar (in terms of gender) to the assigned model. Table 2 additionally shows the descriptive statistics for the same- and cross-gender subgroups.*

7 | DISCUSSION

Previous research on the PF paradigm showed that generating solution attempts prepares students more productively for learning from subsequent instruction compared to the vicarious experience of failure (Kapur, 2014a, 2014b). However, findings by Kapur (2014a, 2014b) also showed that studying examples of students who were engaged in problem solving prior to instruction was more effective compared to receiving instruction first (DI). We hypothesized that the effectiveness of the vicarious experience of failure might depend on the extent to which the *problem-solving-and-failing process* is shown and on whether the observer is able to self-identify with the model observed. We therefore investigated the effects of having access to process information of the PF-model's problem-solving attempts (VF-process), looking only at the finalized outcome of the model's problem solving (VF-outcome), and observing a PF-model with the same or opposite gender. Note that the PF-model was a student from the PF-model group who was engaged in own problem-solving attempts.

Our findings are in accordance with Kapur's (2014a, 2014b) finding that observing examples of failure (VF) can effectively prepare

TABLE 2 Descriptive statistics

Measures		Points/scale	VF same-gender (n = 16) M (SD)	VF cross-gender (n = 17) M (SD)	Total (N = 33) M (SD)
Mathematical ability	Pre-test	1–6	3.09 (1.14)	3.03 (1.17)	3.06 (1.14)
Prior knowledge	Pre-test	0–10	3.50 (0.91)	4.15 (1.23)	3.83 (1.12)
Knowledge gaps	Phase 1	0–5	2.36 (0.69)	2.61 (1.20)	2.49 (0.98)
Solution quantity	Phase 1		4.50 (1.41)	4.24 (1.15)	4.36 (1.27)
Solution quality	Phase 1	0–4	1.88 (0.89)	1.88 (0.93)	1.88 (0.89)
Procedural knowledge	Post-test	0–5	3.75 (1.71)	3.53 (1.33)	3.64 (1.51)
Conceptual knowledge	Post-test	0–7	2.72 (1.70)	1.65 (1.30)	2.17 (1.58)

students for learning from instruction. As stated in Hypothesis 1, students in the DI condition were less able to explain different components of the canonical solution in the post-test (i.e., conceptual knowledge) compared with students in the two VF conditions. Expanding on existing studies on example-based learning, which have already demonstrated the effective use of erroneous examples *after* instruction (e.g., Adams et al., 2014; Heemsoth & Heinze, 2014; McLaren, Adams, & Mayer, 2015) and of worked examples before instruction (Glogger-Frey et al., 2015; Likourezos & Kalyuga, 2017), our results demonstrate that examples of failure can be conducive for learning when studied prior to instruction.

Nevertheless, students who received direct instruction achieved only M = 0.67 (SD = 1.03) points on the conceptual knowledge post-test, and 11 out of 18 students achieved no points at all, calling the effectiveness of the DI control condition into question. Receiving direct instruction might have been too demanding for the students, for instance because the allocated time was not sufficient to impart the content without any preparation or introduction. Therefore, beyond the comparison with a DI condition, a comparison of different example-based conditions (e.g., showing an example of the correct solution as compared to showing examples of failed solution attempts) against each other and with problem-solving as preparatory activity might constitute a better design for further investigating mechanisms underlying the PF approach. An important advantage of example-based control conditions lies in the fact that examples can be designed in accordance with assumed preparatory effects. For instance, in future research, giving students the opportunity to contrasting successful and failing problem-solving behavior, or to compare a failing model's “didactical” responses to failure (e.g., explicitly identifying knowledge gaps) with their own or another students' responses to failing attempts, might provide insights into the importance of failure as a preparatory mechanism.

We argued that the effectiveness of the vicarious experience of failure would depend on the extent to which the problem-solving process is shown. Consequently, we hypothesized that the VF-process condition would prepare students more effectively for the later instruction compared to the condition in which students only looked at the outcome of the model's problem-solving process (Hypothesis 2). However, our data did not support this assumption. While students in the VF-process condition seemed to outperform those in the VF-outcome condition on the conceptual knowledge post-test when looking at the descriptive statistics, this difference was not statistically significant. As our sample size was only sufficient to detect large effects, this may either reflect a power problem, or it might be the case that effective prior knowledge activation does not require students to experience the entire problem-solving process. If students observe and understand the problem-solving attempts, this might be sufficient for them to relate their own prior knowledge to the model's solution attempts, and would help them to elaborate more effectively on the content as taught during the subsequent instruction. Future research with a larger sample size and measures of prior knowledge activation might shed light on this issue.

In addition, we assumed that the awareness of knowledge gaps would explain how effectively students were prepared for the later instruction (Hypothesis 3). When comparing both VF conditions to DI, our data on the students' awareness of knowledge gaps were in line with the students' conceptual knowledge acquisition. However, students who looked only at the outcome of the problem-solving process reported more awareness of knowledge gaps than did students in the VF-process condition, but there were no significant differences in how prepared students in both conditions were to acquire conceptual knowledge from the instruction. If anything, the mean score in the VF-process condition was slightly, though not significantly, higher. One could argue though, that it is the awareness of knowledge gaps as such, rather than the extent of awareness students rated here, that explains the effectiveness of the PF approach. These findings add to research on the PF approach, by showing that not only one's own problem solving can effectively trigger an awareness of knowledge gaps (cf. Loibl & Rummel, 2014a), but also observing examples of failure.

Our findings further suggest that beneficial effects of observing another student's problem-solving attempts depend on the extent to which observing students can self-identify with the model. As stated in Hypothesis 4, students who observed a same-gender model outperformed students who observed a model of the opposite gender on the conceptual knowledge post-test, possibly because they could identify with the model's failure and thus paid more attention to the model's solution attempts. These findings add to the literature on model–observer similarity (e.g., Hoogerheide et al., 2017; Schunk,

1987) and suggest that effects reported in this literature might also apply in the context of PF. In order to manipulate model–observer similarity we used gender as a salient attribute of the model. However, our data do not reveal which specific model characteristics triggered model–observer similarity. Further studies are needed to elucidate which model characteristics indeed trigger the observer's perception of similarity and which cognitive or motivational mechanisms are associated with the perception of model–observer similarity.

Taken together, our study provides important implications for further research on the mechanisms underlying learning from failed solution attempts, particularly when using example-based boundary conditions to investigate preparatory mechanisms of the PF approach. The probably most important implication is that design principles of examples of failure (e.g., problem-solving process or the observer's perception of the model) potentially affect preparatory effects, and example-based control conditions were more effective than the DI control group. However, due to the relatively small sample size and the quasi-experimental design of the study, our results need to be replicated in order to draw more definitive conclusions. Moreover, further studies need to experimentally compare example-based control conditions to a PF condition.

As a first step, our study indicated different ways in which example-based conditions might be designed in order to investigate preparatory effects by means of observing or problem-solving. Another methodological implication for further studies refers to the question of which examples are displayed to the students. In the present study, we displayed all solution attempts generated by the PF-model group to the students in VF. In contrast, Kapur (2014a, 2014b) used the same set of examples for all students in VF. From a research perspective, the yoking-procedure presented here has the advantage that any differences between PF and VF cannot have been caused by differences in the number or kind of solution attempts generated/observed. For instructional practice, however, it would be preferable to show only those examples to the students of which the best possible effects could be expected, but it is still unclear what number and kind of solution attempts would prepare students best for later instruction. An important direction for future research is, therefore, to explore what solution attempts provide the most effective preparation.

If examples of failure effectively prepare students to learn from later instruction, this also has important implications for instructional practice. Firstly, it might relieve students of the cognitive demands resulting from generating their own solution attempts (Van Gog et al., 2019). Secondly, studying examples might be less time-consuming for students than generating their own solution attempts, and this freed-up capacity could be harnessed to stimulate learning by studying examples of failure even further, for instance, by asking students to self-explain the solution attempts (Renkl, 1997). Thirdly, not all students are able to generate the high number of diverse solution attempts necessary in order to be optimally prepared for learning from instruction (Kapur & Bielaczyc, 2012), in which case studying examples of failure might be more effective. Beyond this, a study by Loibl and Rummel (2014a) showed that the effectiveness of the later instruction requires teachers to build on typical student solutions. Teachers might be able to build on examples more effectively, because examples are more predictable than problem-solving attempts generated by students, and examples can be designed more adaptively according to the teacher's intentions. Furthermore, examples might better address typical misconceptions of students, because errors can be highlighted more explicitly, whereas students who fail in problem-solving themselves might have difficulties in recognizing whether their solutions are correct or incorrect. Studying examples of others' failure might also be less frustrating than experiencing one's own failure. In sum, investigating the extent to which examples of failure effectively prepare students for later instruction represents an interesting avenue for further research and instructional practice.

ACKNOWLEDGMENTS

We would like to thank all schools and students for their participation. We also would like to thank our student assistant Charleen Brand for her outstanding commitment to data collection. We would also like to acknowledge Manu Kapur and Katharina Loibl for their interest in discussing our ideas and making materials of their studies available to us.

CONFLICT OF INTEREST

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article.

ORCID

Christian Hartmann https://orcid.org/0000-0003-3109-1104

ENDNOTE

[1] Note that we had originally intended to also compare these conditions with the results of the PF condition, which was used to generate the materials for the two VF conditions. However, as this condition differed too much from the other conditions a priori we did not incorporate it in the analyses and only report the descriptive statistics from this condition (see Table 1 and Figure 3). A comparison of the PF-model group with the remaining conditions (VF-process, VF-outcome and DI) could only be interpreted to a very limited extent, because the PF-model group (a) came from a different school, (b) participated during their free time after class and not during regular lessons, (c) had significantly higher mathematical abilities and prior knowledge and (d) were instructed to think aloud while problem solving in order to produce the modeling examples for VF-process, which could have improved their learning from problem-solving and instruction (see Rittle-Johnson, 2006).

REFERENCES

Adams, D. M., McLaren, B. M., Durkin, K., Mayer, R. E., Rittle-Johnson, B., Isotani, S., & van Velsen, M. (2014). Using erroneous examples to improve mathematics learning with a web-based tutoring system. *Computers in Human Behavior*, *36*, 401–411. https://doi.org/10.1016/j.chb.2014.03.053

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71–81). New York: Academic Press.

Bandura, A., Ross, D., & Ross, S. A. (1963). Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology*, *67*, 601–607. https://doi.org/10.1037/h0045550

Braaksma, M. A. H., Rijlaarsdam, G., & van den Bergh, H. (2002). Observational learning and the effects of model-observer similarity. *Journal of Educational Psychology*, *94*(2), 405–415. https://doi.org/10.1037/0022-0663.94.2.405

Couzijn, M. (1999). Learning to write by observation of writing and reading processes: Effects on learning and transfer. *Learning and Instruction*, *9*, 109–142. https://doi.org/10.1016/S0959-4752(98)00040-1

Darabi, A., Arrington, T. L., & Sayilir, E. (2018). Learning from failure: A meta-analysis of empirical studies. *Educational Technology Research and Development*, *66*(5), 1101–1118. https://doi.org/10.1007/s11423-018-9579-9

Glogger-Frey, I., Fleischer, C., Grüny, L., & Renkl, A. (2015). Inventing a solution and studying worked solution prepare differently for learning from direct instruction. *Learning and Instruction*, *39*, 72–87. https://doi.org/10.1016/j.learninstruc.2015.05.001

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & van den Bergh, H. (2013). The effect of observational learning on students' performance, processes, and motivation in two creative domains. *British Journal of Educational Psychology*, *83*, 31–45. https://doi.org/10.1111/j.2044-8279.2011.02052.x

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. https://doi.org/10.1007/s11251-013-9302-5

Hoogerheide, V., Loyens, S. M. M., Jadi, F., Vrins, A., & Van Gog, T. (2017). Testing the model-observer similarity hypothesis with text-based worked examples. *An International Journal of Experimental Educational Psychology*, *37*(2), 112–127. https://doi.org/10.1080/01443410.2015.1109609

Hoogerheide, V., Loyens, S. M. M., & Van Gog, T. (2016). Learning from video modeling examples: Does gender matter? *Instructional Science*, *44*(1), 69–86. https://doi.org/10.1007/s11251-015-9360-y

Kapur, M. (2012). Productive failure in learning the concept of variance. *Instructional Science*, *40*(4), 651–672. https://doi.org/10.1007/s11251-012-9209-6

Kapur, M. (2014a). Productive failure in learning math. *Cognitive Science*, *38*, 1008–1022. https://doi.org/10.1111/cogs.12107

Kapur, M. (2014b). Comparing learning from productive failure and vicarious failure. *The Journal of the Learning Science*, *23*, 651–677. https://doi.org/10.1080/10508406.2013.819000

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *The Journal of the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Likourezos, V., & Kalyuga, S. (2017). Instruction-first and problem-solving-first approaches: Alternative pathways to learning complex tasks. *Instructional Science*, *45*, 195–219. https://doi.org/10.1007/s11251-016-9399-4

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. https://doi.org/10.1007/s10648-016-9379-x

Loibl, K., & Rummel, N. (2014a). Knowing what you don't know makes failure productive. *Learning and Instruction*, *42*(3), 305–326. https://doi.org/10.1016/j.learninstruc.2014.08.004

Loibl, K., & Rummel, N. (2014b). The impact of guidance during problem solving prior to instruction on students' inventions and learning outcomes. *Instructional Science*, *42*(3), 305–326. https://doi.org/10.1007/s11251-013-9282-5

Loibl, K., & Rummel, N. (2015). Productive failure as strategy against the double curse of incompetence. *Learning: Research and Practice*, *1*(2), 113–121. https://doi.org/10.1080/23735082.2015.1071231

McLaren, B. M., Adams, D. M., & Mayer, R. E. (2015). Delayed learning effects with erroneous examples: A study of learning decimals with web-based tutor. *International Journal of Artificial Intelligence in Education*, *25*(4), 520–542. https://doi.org/10.1007/s40593-015-0064-x

Renkl, A. (1997). Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. *European Journal of Psychology of Education*, *14*(4), 477–488. https://doi.org/10.1007/BF03172974

Rittle-Johnson, B. (2006). Promoting transfer: Effects of self-explanation and direct instruction. *Child Development*, *77*(1), 1–15. https://doi.org/10.1111/j.1467-8624.2006.00852.x

Schunk, D. H. (1987). Peer models and Children's behavioral change. *Review of Educational Research*, *57*(2), 149–174. https://doi.org/10.3102/00346543057002149

Tsovaltzi, D., McLaren, B. M., Melis, E., & Meyer, A.-K. (2012). Erroneous examples: Effects on learning fractions in a web-based setting. *International Journal of Technology Enhanced Learning*, *4*(3/4), 191–230. https://doi.org/10.1504/IJTEL.2012.051583

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review*, *22*, 155–174. https://doi.org/10.1007/s10648-010-9134-7

Van Gog, T., Rummel, N., & Renkl, A. (2019). Learning how to solve problems by studying examples. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 183–208). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108235631.009

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Hartmann C, van Gog T, Rummel N. Do examples of failure effectively prepare students for learning from subsequent instruction? *Appl Cognit Psychol*. 2020;34:879–889. https://doi.org/10.1002/acp.3651