



Preparatory effects of problem solving versus studying examples prior to instruction

Christian Hartmann¹ · Tamara van Gog² · Nikol Rummel¹

Received: 1 March 2019 / Accepted: 20 November 2020 / Published online: 4 January 2021
© The Author(s) 2021

Abstract

The Productive Failure (PF) approach prompts students to attempt to solve a problem prior to instruction – at which point they typically fail. Yet, research on PF shows that students who are involved in problem solving prior to instruction gain more conceptual knowledge from the subsequent instruction compared to students who receive the instruction first. So far, there is no conclusive evidence, however, that the beneficial effects of PF are explained by the attempt to generate *one's own* solutions prior to instruction. The literature on example-based learning suggests that observing someone else engaging in problem-solving attempts may be an equally effective means to prepare students for instruction. In an experimental study, we compared a PF condition, in which students were actively engaged in problem solving prior to instruction, to two example conditions, in which students either observed the complete *problem-solving-and-failing* process of another student engaging in PF or looked at the outcome of this process (i.e., another student's failed solution attempts). Rather than worked examples of the correct solution procedure, the students observed examples of failed solution attempts. We found that students' own problem solving was not superior to the two example conditions. In fact, students who observed the complete PF process even outperformed students who engaged in PF themselves. Additional analyses revealed that the students' prior knowledge moderated this effect: While students who observed the complete PF process were able to take advantage of their prior knowledge to gain more conceptual knowledge from the subsequent instruction, prior knowledge did not affect students' post-test performance in the PF condition.

Keywords Productive failure · Vicarious failure · Example-based learning · Observational learning · Mathematics · Conceptual knowledge acquisition

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11251-020-09528-z>) contains supplementary material, which is available to authorized users.

✉ Christian Hartmann
christian.hartmann@rub.de

¹ Institute of Educational Research, Ruhr University Bochum, Universitätsstraße 150, 44801 Bochum, Germany

² Department of Education, Utrecht University, P.O. Box 80140, 3508 TC Utrecht, The Netherlands

Introduction

Research on Productive Failure (PF; Kapur 2012) has found strong support for the beneficial preparatory effects of problem solving prior to instruction, especially in mathematics and science education (Darabi et al. 2018). Unlike problem-based learning approaches, where students learn while working on an ill-defined problem which can have multiple solution paths or even multiple ‘correct’ solutions and instruction is only provided if needed (for an overview: Lu et al. 2014), the PF approach uses complex (i.e., in the sense that there are many interacting information elements; Ashman et al. 2019) and challenging but not ill-defined problems (i.e., once the prerequisite conceptual knowledge has been acquired, the solution procedure is straightforward) and foresees students attempting to identify some of the underlying components of the correct solution using cues from the given information on the problem. Attempting (and failing) to solve such a problem prior to instruction seems to make students more receptive to the subsequent instruction. Students who learn with this PF approach demonstrate greater improvement in their conceptual knowledge acquisition compared to students learning with a direct-instruction (DI) approach in which instruction is provided first, followed by problem solving (Kapur 2012). While students who engage in PF do not gain more procedural knowledge (i.e., knowledge about how to apply a formula) compared to DI, they do perform better on later tests of conceptual knowledge about formula components and their functions and relations (for an overview: Loibl et al. 2017). Even if the student does not yet know the concept required to solve a mathematical problem, the attempt to solve the problem is assumed to trigger the mechanisms of prior knowledge activation and awareness of knowledge gaps, which *prepare* the student to benefit from subsequent instruction (Loibl et al. 2017). These mechanisms are expected to support students in cognitively processing, organizing, and integrating the information on the new concept that is provided to them during the subsequent instruction more effectively (i.e., deep feature recognition) as compared to the direct-instruction approach.

However, despite convincing evidence that the PF approach is more effective than the direct-instruction approach for students’ conceptual knowledge acquisition (and some indirect evidence that the above-mentioned mechanisms play a role; Loibl et al. 2017), it remains unclear whether students have to engage in generating *their own* solution attempts prior to instruction, or whether other ways to trigger these mechanisms, for instance *observing someone else* engaging in problem-solving attempts, would be as effective. Compared to generating one’s own solution attempts, studying examples of other students attempting (and failing) to solve a problem, would be an interesting addition for educational practice: It is easy to implement in online as well as classroom learning environments, it may be less frustrating and less cognitively demanding than attempting (and failing) to solve a problem oneself without a loss of effectiveness. So far, PF studies assume challenging students is positive, but studies have not yet addressed motivational effects of PF or VF, that is, whether students feel frustrated, cognitively challenged or even overwhelmed while problem solving. Therefore, observing examples of failure might be less challenging and still prepare students effectively for later instruction. Moreover, it may benefit all students and not just those who are able to generate a high variety of solution attempts. That is, prior research showed that the PF-effect is strongest for students who manage to generate a high variety of solution attempts (Kapur 2014a; Kapur and Bielaczyc 2012), but there is substantial variety among students in the number and quality of solution attempts they manage to generate (cf. Wiedmann et al. 2012). If studying examples of others’ failed solution attempts can also effectively prepare students for instruction, those students who are

not yet able to generate a high variety of solution attempts would particularly benefit from studying a set of examples representing a sufficient variety in solution attempts.

Example-based learning has been shown to be effective particularly for acquiring (procedural) knowledge about how to correctly solve a problem in a variety of domains (Van Gog et al. 2019). However, the question remains whether *observing someone else failing to solve a problem* would also be productive for one's own learning from subsequent instruction. There are already some indications in research on the PF paradigm that studying the *outcomes* of failed solution attempts of other students prior to instruction is more effective than DI (though not as effective as PF). Kapur (2014a,b) compared a PF condition to a 'Vicarious Failure' (VF) condition in which students were instructed to evaluate failed solution attempts created by other students prior to receiving instruction (i.e., none of the examples displayed the correct solution). Students in both the PF and the VF condition gained more conceptual knowledge from the subsequent instruction than students in a direct-instruction condition, although the results also revealed that students in the PF condition outperformed those in the VF condition on a conceptual knowledge post-test (Kapur 2014a). The fact that VF outperformed DI suggests that observing another student engaging in PF may trigger similar beneficial mechanisms as engaging in PF oneself – at least to some extent, as VF was still less effective than PF.

One explanation for the lower effectiveness of VF as compared to PF in Kapur's study might lie in the fact that the VF students only saw the *outcomes* of the failed solution attempt of the other student, and not the complete problem-solving process. Perhaps some essential parts of the problem-solving process in PF, which might be relevant for the preparatory effects of PF for learning from the subsequent instruction, were withheld from the VF students (e.g., the other students' intentions behind the solution attempts, as well as reflections or conclusions). That is, the examples did not show how the PF model actively explored the not-yet-known concept by activating prior knowledge and becoming aware of knowledge gaps. This process information might, however, be important for the observing students in order to activate their own prior knowledge and become aware of their own knowledge gaps. Even if the activation of prior knowledge could also be an implicit cognitive mechanism (i.e., an unconscious process from which abstract knowledge results; for an overview: Seger 1994), existing theories on the PF approach have emphasized explicit problem-solving processes as preparatory mechanisms: the solution attempts as a proxy for prior knowledge activation and the awareness of knowledge gaps. Therefore, having full access to the problem-solving process might be an essential precondition for an effective preparation for the subsequent instruction and thus for learning. Against this background, the aim of the experimental study we conducted was to compare the effectiveness of students attempting to generate their own solutions prior to instruction to (1) students observing others attempting (and failing) to generate solutions (i.e., observing the 'failure' process), and (2) students looking at another student's failed solution attempts (i.e., looking only at the outcome of the process, without information about the problem-solving process).

Problem solving prior to instruction: doing versus observing

In this section we first describe the assumed preparatory effects of problem-solving prior to instruction (i.e., prior knowledge activation and awareness of knowledge gaps). Then we review relevant literature on example-based learning and discuss, in light of this literature,

why one might expect that observing another student engaging in problem-solving attempts might trigger similar beneficial preparatory effects as problem solving prior to instruction.

Preparatory effects of problem solving

When students engage in problem-solving attempts before receiving instruction, they might *activate relevant prior knowledge* of parts of the solution (i.e., incomplete schemas), which might help them process the information provided during the subsequent instruction more deeply (Loibl et al. 2017). The activation of preexisting schemas during their failed solution attempts might help students integrate the new conceptual knowledge provided to them during the instruction more effectively and with less mental effort (Sweller et al. 1998). In contrast, the DI approach requires students to activate their prior knowledge while simultaneously processing and integrating the new information given in the instruction.

Loibl and Rummel (2014a) additionally assume that students who attempt to solve a problem prior to instruction, and fail, might develop an *awareness of their knowledge gaps*. If students know that (or even why) a solution attempt *cannot* be the canonical solution, they may be able to fill their gaps more effectively during the subsequent instruction. Consequently, being taught how to overcome their misconceptions during the subsequent instruction becomes more effective due to the students' initial awareness of their knowledge gaps. Even if the students are, at this point, not yet able to specify what components of the correct (or canonical) solution are missing, they are aware that their solution attempts lead to erroneous results. Due to this awareness, they might also become more curious about the not-yet-known concept, and thus more motivated to learn from the instruction.

As we will argue in the following two sections, in light of the literature on example-based learning, one might expect that *observing* another student engaging in *problem-solving attempts* or *studying another student's solutions* could also be effective in preparing students to learn from subsequent instruction. In research on example-based learning, students are usually presented with examples that demonstrate the correct application of the problem-solving procedure leading to a canonical solution (i.e., worked examples), and are then asked to solve similar problems by themselves (e.g., Sweller and Cooper 1985). This is very different from what an 'example-based PF condition' would look like, in which students observe others failing at solving a problem and thereby creating incorrect or incomplete solution ideas (for a discussion, see Kalyuga and Singh 2016). Moreover, in most example-based learning studies the main outcome of interest is procedural knowledge, whereas in PF it is conceptual knowledge. Yet, there is some evidence from research on observational and example-based learning that observing another student who is attempting (and failing) to solve problems may have similar effects as attempting (and failing) to solve problems oneself (i.e., may also lead to prior knowledge activation, awareness of knowledge gaps or similar beneficial mechanisms). For instance, studies on erroneous examples show that observing another student's errors can be effective for conceptual learning (e.g., Tsovaltzi et al. 2012; Durkin and Rittle-Johnson 2012) and transfer (Große and Renkl 2007), although erroneous examples usually have not been presented to learners *prior to* instruction. Existing studies indicate that, compared to being engaged in problem solving by oneself, studying (correct) worked examples prior to instruction on the one hand, and studying erroneous examples after instruction on the other hand, can also be effective for conceptual learning. Probably because these learning activities also support students in activating prior knowledge, becoming aware of their knowledge gaps

or similarly beneficial mechanisms (e.g., deep feature recognition). After discussing relevant studies from these two strands of research in more detail, we will highlight an important gap in the research that we address in the present study: Studying examples of failing solution attempts *prior to* instruction (hereon referred to as ‘examples of failure’).

Preparatory effects of studying worked examples

Research on example-based learning provides convincing evidence that observing a demonstration of how to solve a problem can be more effective and efficient (i.e., requiring less time and effort) for learning than engaging in problem solving oneself (for reviews, see Van Gog and Rummel 2010; Van Gog et al. 2019). In addition, research on example-based learning has shown that it is particularly beneficial when students attempt to self-explain the principles underlying the problem-solving steps, as this helps learners to understand the deep features of a worked example (Renkl 2002; Renkl and Eitel 2019). While in most studies worked examples have been provided *after instruction* (if instruction was provided at all), there is at least some evidence that studying worked examples *prior to instruction* can be as effective for learning as problem solving prior to instruction. For instance, Likourezos and Kalyuga (2017) compared a condition in which students invented solutions without guidance (own problem solving) with two example conditions in which students either received partial guidance (some parts of the canonical solution) or full guidance (complete correct worked solution) *prior to instruction* on the canonical solution. The results did not reveal differences in a delayed transfer post-test between students who studied examples prior to instruction and those who previously solved the problem on their own. These findings are in line with two experimental studies by Glogger-Frey et al. (2015), in which the authors compared students who invented solutions for a problem prior to instruction to students who studied and self-explained a correctly worked-out solution to the same problem prior to instruction (cf. the full guidance condition in Likourezos and Kalyuga 2017). Students in the worked example group outperformed those in the invention condition in a delayed transfer post-test after instruction (Glogger-Frey et al. 2015). The findings of Glogger-Frey et al. (2015) and Likourezos and Kalyuga (2017) support the assumption that studying examples prior to instruction can effectively prepare students for learning from the subsequent instruction. Glogger-Frey et al. (2015) argue that studying worked examples prior to instruction prepares students for learning from a later instruction, because by self-explaining deep features of the presented worked examples they already acquired well-organized knowledge about the learning content, and this helped them to better connect and integrate the content of the later instruction.

However, the examples used in the above-mentioned studies were not examples of failure, but correct ones and thus allowed students to study how to apply essential parts of the canonical solution correctly prior to instruction. This is very different from what an ‘example-based PF condition’ would look like, in which students observe others failing at solving a problem, meaning they observe incorrect and incomplete solution attempts (i.e., not covering all elements of the canonical solution). As mentioned earlier, however, there is also some evidence that studying erroneous examples can be effective for conceptual learning, transfer, and more effective than being engaged in problem solving oneself (e.g., McLaren et al. 2015).

Preparatory effects of studying erroneous examples

In contrast to (correct) worked examples, erroneous examples ask students to *detect and correct errors* in problem-solving attempts. For instance, McLaren et al. (2015) replicated the findings of Adams et al. (2014) that students who studied errors made by fictitious students within a computer-supported learning environment, learned more than students who solved a problem by themselves with additional feedback. Adams et al. (2014) stressed that studying erroneous examples fosters deeper cognitive processing aimed at organizing the learning material and relating it to students' own prior knowledge.

Furthermore, there are some indications that studying erroneous examples (after instruction) also triggers an awareness of knowledge gaps, which is relevant in light of the assumption that the awareness of knowledge gaps is one of the cognitive mechanisms underlying the effectiveness of PF (Loibl et al. 2017). Adams et al. (2014), for instance, argued that studying erroneous examples enables students to be more aware of their acquired knowledge, as the authors found that students who studied erroneous content prior to the post-test were able to more accurately judge their own (correct and incorrect) post-test performance. In other words, studying erroneous examples enabled students to assess more precisely what they did and did not know. In line with this finding, a study by Heemsoth and Heinze (2014) revealed that students who reflected on erroneous examples acquired more negative knowledge (i.e., knowledge about why a certain solution does not work) than students who reflected on correct examples.

It should be noted, however, that the erroneous examples in existing studies (e.g., Adams et al. 2014) were studied *after instruction*. Thus, participants had already received instruction before studying the erroneous examples or had access to the canonical solution while studying the erroneous examples. Consequently, they already had knowledge about the correct problem solution. It therefore remains unclear whether studying examples of failed solution attempts could also support students' conceptual knowledge acquisition if they did not receive relevant information about how to identify and overcome their misconceptions until the *subsequent instruction*, as would be the case in the context of an example-based PF approach (i.e., examples of failure). In other words, it remains unclear whether studying examples of failure would have similar preparatory effects and help students to gain more conceptual knowledge from subsequent instruction as engaging in PF. Building on studies by Kapur (2014a,b), this is the research gap we address in the present study.

Research questions and hypotheses

As discussed in the previous sections, existing research supports the beneficial effects of studying worked examples (i.e., of [parts of] the canonical solution) prior to instruction and of studying erroneous examples after instruction, but it is as yet unclear whether studying *examples of failure prior to receiving instruction* can be as conducive to learning as engaging in PF. Kapur (2014a, b) compared a PF condition to a 'Vicarious Failure' (VF) condition in which students were instructed to evaluate failed solutions generated by other students prior to receiving instruction. Students in both the PF and the VF condition gained more conceptual knowledge from the subsequent instruction than students in a DI condition, although the results also revealed that students in the PF condition outperformed those in the VF condition on a conceptual knowledge post-test (Kapur 2014a).

We argue that an explanation for the lower effectiveness of VF than PF in Kapur's study might lie in the fact that the VF students only saw the outcomes of the failed solution attempt of the other student, and not how the PF model actively explored the not-yet-known concept. This process information, however, might be important for the observing students in order to relate the model's solution attempts to their own prior knowledge (i.e., prior knowledge activation). Having access, for instance, to the model's intentions behind the solution attempts, as well as reflections or conclusions, might help students in the VF-process condition to become aware of their own knowledge gaps as well.

In the present study, accordingly, we aimed to extend the findings of Kapur (2014a, b) by experimentally comparing a PF condition in which students are actively involved in their own problem solving prior to instruction to two example conditions, in which students either observed the complete problem-solving process of another student engaging in PF (VF-process) or looked at the solutions (VF-outcome; cf. Kapur (2014a, b). We also assessed the students' self-reported awareness of knowledge gaps to further explore preparatory effects. Moreover, we evaluated the number and quality of the generated solution attempts in the PF condition as a proxy of students' prior knowledge activation (Loibl and Rummel 2014b). As Kapur (2014a) found that PF and VF were more effective than DI, and Hartmann et al. (2020) replicated this result in a quasi-experimental study, we did not implement a DI condition here.

Hypotheses

As previous research on PF reported beneficial effects on conceptual knowledge acquisition, while procedural knowledge acquisition was typically not affected (Loibl and Rummel 2014b; Kapur 2014a, b), we expected no significant differences between the conditions with regard to procedural knowledge. We formulated the following hypotheses:

Hypothesis 1

With regard to conceptual knowledge acquisition, as outlined above, we hypothesized that observing the problem-solving process prior to instruction (VF-process) would be as effective in terms of preparing students for learning from the subsequent instruction as engaging in problem solving (PF).

Hypothesis 2

We hypothesized that students in the PF and VF-process conditions would show more conceptual knowledge on the post-test than students who only looked at the outcome of the problem-solving process (VF-outcome).

Further exploration of preparatory effects

Despite indirect evidence from previous studies that the awareness of knowledge gaps and activation of prior knowledge might explain the effectiveness of the PF approach (Loibl et al. 2017), the effect of these mechanisms remains unclear. Therefore, in addition to our main hypotheses on conceptual knowledge acquisition, we also explored the impact on awareness of knowledge gaps and prior knowledge activation.

We expected the results regarding students' self-reported knowledge gaps to be in line with Hypotheses 1 and 2, because we assume that observing the problem-solving process (VF-process) would trigger similar mechanisms as PF, which would prepare students to benefit from subsequent instruction. Therefore, students in the PF and VF-process conditions should be similarly aware of their knowledge gaps, and compared to students in the VF-outcome condition, PF and VF-process should be even more aware. Accordingly, we also assumed a positive correlation between students' awareness of knowledge gaps and their performance on the conceptual knowledge post-test.

In addition, we evaluated the number and quality of the generated solution attempts in the PF condition as a proxy of students' prior knowledge activation. We expected that the greater the amount of relevant prior knowledge activated by students (i.e., quantity of solution attempts), the more they would gain from the subsequent instruction. Kapur (2012, 2014b) reported a positive correlation between the number of solution attempts generated during PF and conceptual knowledge acquisition, but Loibl and Rummel (2014b) did not confirm this finding. Although the results of previous studies are mixed, we assumed a positive correlation between the number of solution attempts generated during PF and conceptual knowledge acquisition. It should be noted that we only tested this hypothesis for the PF condition; the students in both VF conditions studied exactly the same solution attempts, as was the case in Kapur's studies, and there was no variance in the number and quality of solutions within our VF conditions.

Methods

Participants and design

We conducted our experiment at five secondary schools (nine classes) in Germany during regular mathematics lessons. The sample consisted of $N=198$ students (98 female and 100 male; age: $M=16.10$, $SD=0.78$), who were randomly assigned to one of the three conditions: PF ($n=63$), VF-Process ($n=65$), or VF-Outcome ($n=70$). Twenty-five students had to be removed from the dataset (PF: $n=8$, VF-process: $n=9$, VF-outcome: $n=8$) because: (a) they already knew the concept to be learned (PF: $n=0$, VF-process: $n=0$, VF-outcome: $n=3$), (b) had exceptionally low mathematics grades (i.e., were outliers; PF: $n=3$, VF-process: $n=1$; VF-outcome: $n=0$), or (c) they did not attempt to answer *any* questions on the post-test (PF: $n=5$, VF-process: $n=8$, VF-outcome: $n=5$), which strongly suggests they lacked the motivation to seriously engage in the post-test (and as such, a zero score on the posttest in these cases would not reflect/measure a lack of learning). The final sample thus comprised $N=173$ (87 female and 86 male; age: $M=16.06$, $SD=0.77$) students, which is sufficient to detect a medium-sized effect ($f=0.23$, $1-\beta=0.80$; G-Power Analysis).

Materials

Problem-solving task

The mathematical problem-solving task used in the initial learning phase was adopted from Kapur (2012), and is the most frequently employed task in PF studies (e.g., Loibl and Rummel 2014a, b; Kapur 2014a, b). The task targets the concept of variance (standard or mean absolute deviation) using a cover story about soccer: Students need

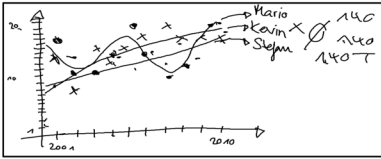
to find the most consistent soccer player by analyzing a list of the number of goals that three soccer players scored over a 10-year period (see Loibl and Rummel 2014a). Students in the PF condition usually generate a number of solution attempts that differ in how many of the following components of the canonical solution they cover: (1) sum up deviations for all data points to obtain a precise result, (2) take absolute or squared deviations (i.e., positive values) to prevent positive and negative deviations from canceling each other out, (3) take deviations from a fixed reference point (the mean) to avoid sequence effects, and (4) divide by the number of data points to account for sample size. As students have not yet received instruction on the concept of variance and as shown by previous studies (e.g., Kapur 2014a; Loibl and Rummel 2014a), the students usually do not come up with the canonical solution during the problem-solving phase. A process analysis of 24 PF-students' think-aloud protocols from a previous sample of the study by Hartmann et al. (2020), in which students worked on the same problem-solving task as used in the present study, showed that 91.7% of the students noticed limitations of their solution attempts (Brand et al. 2018). It can be assumed, accordingly, that if students attempt to solve the problem-solving task used in this study, they usually are aware that they did not correctly solve the problem. Furthermore, the problem-solving task was designed such that the students cannot be sure if their solution attempts were correct. For instance, in line with design principles of the PF approach (Kapur and Bielaczyc 2012), the task instruction prompted students to find more than just one problem solution. The typically produced problem solutions also lead to very different results which also prevent students from being entirely certain about the correctness of their solution attempts. The problem-solving task can be found in the supplementary materials (Appendix A).

Examples used in the VF conditions

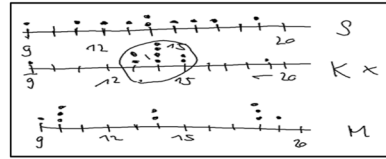
The examples used in the VF conditions were obtained from an initial quasi-experimental study (Hartmann et al. 2020), in which 24 students attempted to find solutions for the above-mentioned mathematical problem within a PF condition. The students produced their solutions on tablet PCs while thinking out loud, and this process was audio- and video-recorded. Based on the recordings, one student's solutions were selected to serve as examples in the present study (see Fig. 1). The selected student demonstrated clear handwriting on the tablet, generated six solutions, and did not find the canonical solution (the highest-quality solution covered two out of the four canonical components). As such, this model student represents a typical student comparable with previous studies (e.g., Loibl and Rummel 2014b).

In both example conditions (VF-process and VF-outcome), we tried to prevent a potential effect of model-observer similarity (see Hartmann et al. 2020; for a discussion of MOS, see Hoogerheide et al. 2016) by controlling the gender match between model and observer. To this aim, we also recorded the problem-solving process of the originally female model with a male voice. As participants in the VF-outcome condition did not hear the model's voice, we introduced fictional female or male names, and displayed them in accordance with the gender of the observer. Therefore, students in both example conditions worked with same-gender models. In order to trigger and thus control for MOS, the students were also told that the model had a similar level of ability to themselves.

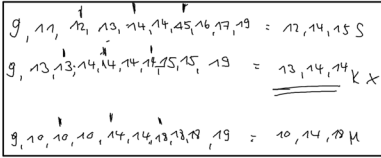
Solution 1 – Trend lines & Sum of data points



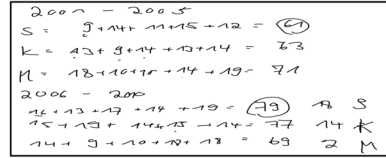
Solution 2 – Dot Diagram



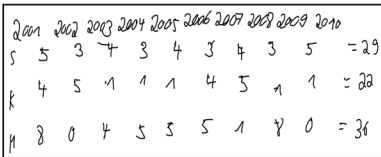
Solution 3 – Median & Distribution



Solution 4 – Year-on-Year Differences (periods)



Solution 5 – Year-on-Year Differences



Solution 6 – Differences sorted in ascending order

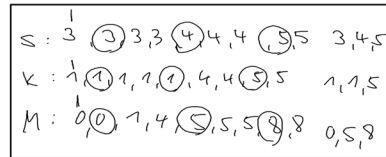


Fig. 1 Illustration of the six examples used in the VF conditions

Mathematical ability and prior knowledge (Pre-test)

We assessed students’ mathematical ability and prior knowledge as potential covariates to take into account in the analysis. *Mathematical ability* was assessed by asking students to state their last two grades in mathematics. Thus, mathematical ability indicates a rather global skill to solve mathematical problems as evidenced by past academic performance. Students’ *prior knowledge* was assessed with a pre-test consisting of five mathematical tasks (Cronbach’s alpha=0.402). As we did not expect the students to have knowledge about the concept of variance before our experiment, the prior knowledge test assessed relevant prerequisite knowledge rather than specific knowledge about the concept of variance, although we checked whether the students already knew the canonical formula prior to the experimental phase. The pre-test was identical to the one used by Loibl and Rummel (2014a, b) as well as Hartmann et al. (2020) and required students to interpret and draw graphs (3 points), demonstrate their knowledge of descriptive statistics (mean and range, 3 points), and to draw and interpret a boxplot (4 points).

Quantity and quality of solution attempts (Process data, PF)

We evaluated the quantity and quality of generated student solutions in the PF condition as a proxy for prior knowledge activation using the coding scheme by Loibl and Rummel (2014b). Each solution received a score ranging from 0 (none of the canonical solution components included) to 4 (all of the canonical solution components included). To assess the overall quality of the PF students’ problem solving, we used their score on the

solution with the highest number of components included. To assess the quantity of student solutions, we counted the number of different solutions, regardless of their quality. For instance, the model in our VF conditions generated six solutions (quantity) with two components of the canonical solution included (quality).

Self-reported awareness of knowledge gaps

Students' *awareness of knowledge gaps* was assessed with items adopted from Loibl and Rummel (2014a) and Glogger-Frey et al. (2015). Seven items (Cronbach's $\alpha=0.806$) rated on a 6-point Likert scale were used to measure the extent to which students perceived their own knowledge after problem solving (or observing the model or looking at the model solutions, respectively) as sufficient. Some of the items asked about the generated or observed solution attempts more directly (e.g., 'the solutions seem to be incomplete'). All items can be found in the supplementary materials (Appendix B).

Conceptual and procedural knowledge (Post-test)

The post-test, which assessed *conceptual* and *procedural knowledge*, was identical to that used by Loibil and Rummel (2014a, b) and Hartmann et al. (2020). The four conceptual knowledge items required students to explain graphical representations of the canonical formula (2 points), to identify and explain errors of typical student solutions by relating them to essential components of the canonical solutions (3 points), and use their conceptual understanding of 'consistency' to sort data sets according to their distribution (2 points). The three procedural knowledge items required students to apply the canonical problem-solving procedure (i.e., mean absolute deviation) to problems as shown during the instruction (5 points). For reliability analysis, we calculated Cronbach's α to assess the internal consistency of the whole post-test and of the two subscales. The internal consistency of the seven items was Cronbach's $\alpha=0.663$ (separately by subscale, Cronbach's α was 0.437 for the conceptual and 0.593 for the procedural items). As our hypotheses mainly concerned the conceptual knowledge post-test, a second rater coded all of the conceptual knowledge post-tests of all students. Inter-rater reliability assessed by the *ICC* (random, absolute) was high for the entire scale, $ICC=0.946$, 95% *CI* [0.865, 0.972], as well as for each item.

In addition to the main variables relating to our hypotheses, we also collected self-report data on perceived competence, curiosity, intrinsic motivation, cognitive engagement, epistemic beliefs about mathematics, self-efficacy, need for cognition, and perceived model-observer similarity. As we did not have any hypotheses about these exploratory measures, we did not consider them in our main analyses.

Procedure

The study comprised of two experimental phases (for an overview, see Fig. 2). After completing the pre-test (duration of 20 min), students worked on the mathematical problem for 45 min. The experimental variation was implemented in this phase. The students remained in their classes as usual, and worked on tablet PCs with headphones on. Before students started to work on the mathematical problem, we randomly assigned each student to one of the three experimental conditions.

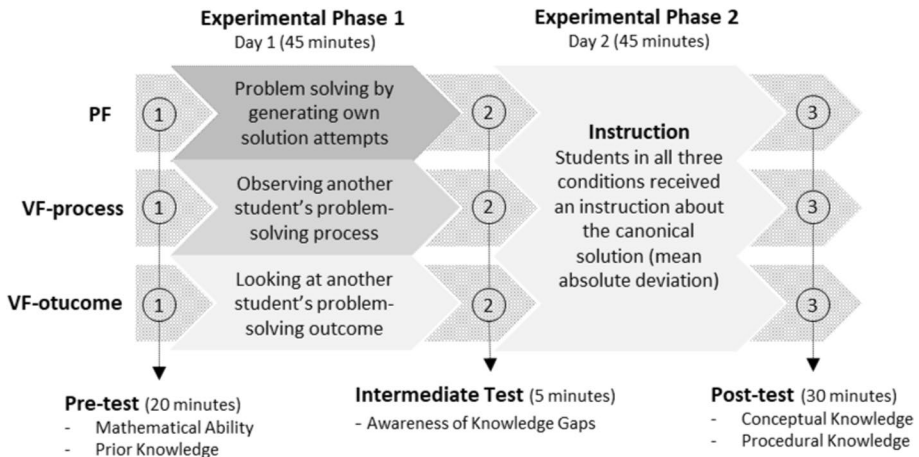


Fig. 2 Experimental procedure

At the beginning of the first experimental phase, we prompted students in the PF condition to find as many solutions for the mathematical problem as they could; thus, the PF students invented their own solution attempts. All of their solution attempts were recorded on the tablet PCs, including any notes made. Students in the two VF conditions did not invent their own solution attempts, but rather observed solutions generated by another student, displayed on the tablet PCs. In the introduction to the experimental phase, we told the students in the two VF conditions that they were going to spend 45 min observing another student who was attempting to solve a problem. Even if the introduction stressed that the other student was *attempting* to solve the problem, in accordance with the introduction used by Kapur (2014a, b) we did not explicitly highlight that the shown solutions were erroneous. The two VF conditions differed regarding the information to which students had access, but the model's solutions were identical for both conditions. Students in the example condition with process information (VF-process) observed the model generating his/her solutions. They watched a video showing the model's problem-solving steps and could hear the model's voice throughout the process. Students in the VF-outcome condition were shown the same solutions as students in the VF-process condition, but were only shown pictures of the final state of the model's solutions, and did not receive the audio and video. Each solution was shown for the same amount of time that the model had needed to generate the solution (i.e., same duration as the VF-process condition). Students did not receive any guidance or instruction on the target concept or concerning relevant problem-solving strategies during this phase in any of the conditions.

After the experimental phase, in the next mathematics lesson within one week, students in all three conditions received 45 min of instruction about the canonical solution (i.e., mean absolute deviation) from the experimenter. This instruction was the same as that used in the study by Hartmann et al. (2020), and contrasted typical student solutions to the canonical solution (cf. Loibl and Rummel 2014a). The instruction was held as the students' regular mathematics lessons and the students remained in their classes as usual. The experimenter introduced the problem-solving task and typical (failed) solution attempts (as generated or observed in the previous phase). For each solution attempt, the experimenter highlighted the advantages and disadvantages. The solution approaches were presented in

ascending quality until the canonical solution was introduced. At the end of the instruction, the experimenter summarized all failed solution attempts and demonstrated why the canonical solution better solves the problem-solving task. Finally, students completed a 30-min post-test requiring them to apply (procedural knowledge) and explain (conceptual knowledge) the canonical solution to tasks similar to the mathematical problem used in the experimental phase.

Results

Mathematical ability and prior knowledge

Table 1 displays the descriptive statistics for students' mathematical ability and their prior knowledge. As revealed by two separate ANOVAs, the three conditions did not differ significantly in mathematical ability, $F(2, 170) = 2.54$, $p = 0.082$, $\eta_p^2 = 0.03$ or prior knowledge, $F(2, 170) = 0.76$, $p = 0.469$, $\eta_p^2 = 0.01$.

Conceptual knowledge

Table 2 shows the descriptive statistics for conceptual and procedural knowledge acquisition as well as the overall score aggregated across both subscales.

We tested our hypotheses regarding the students' conceptual knowledge acquisition using an ANCOVA, defining two a priori contrasts. The first contrast tests Hypothesis 1 that observing the problem-solving process prior to instruction (VF-process: weight of -1) is as effective for preparing students for the subsequent instruction as engaging in problem solving oneself (PF: weight of 1 ; VF-outcome was assigned a weight of 0). The second contrast tests Hypothesis 2 that students in both the PF (weight of 1) and VF-process (weight of 1) condition gain more conceptual knowledge than students who look at the outcome of the problem-solving process (VF-outcome; weight of -2).

Given that both mathematical ability, $r(173) = 0.36$, $p < 0.001$, and prior knowledge, $r(173) = 0.31$, $p < 0.001$ correlated positively with conceptual knowledge acquisition, these measures were used as covariates. Following suggestions by Hayes (2013), we mean-centered both covariates. The ANCOVA revealed a significant effect of prior knowledge, $F(1, 166) = 14.43$, $p < 0.001$, $\eta_p^2 = 0.08$, and mathematical ability, $F(1, 166) = 14.03$, $p < 0.001$, $\eta_p^2 = 0.08$.

With regard to our first hypothesis that VF-process would be as effective in terms of preparing students for the subsequent instruction as PF, the contrast showed that VF-process

Table 1 Descriptive statistics (pre-test)

Pre-test	Points /scale	PF <i>n</i> = 55 <i>M</i> (<i>SD</i>)	VF-process <i>n</i> = 56 <i>M</i> (<i>SD</i>)	VF-outcome <i>n</i> = 62 <i>M</i> (<i>SD</i>)	TOTAL <i>N</i> = 173 <i>M</i> (<i>SD</i>)
Mathematical ability ^a	1–6	3.92 (1.01)	4.30 (0.94)	4.24 (0.90)	4.16 (0.96)
Prior knowledge	0–10	5.27 (2.01)	4.98 (1.74)	5.40 (1.85)	5.22 (1.87)

^aIn Germany, '1' represents the best grade while '6' represents the worst grade. To facilitate interpretation, grades were reverse-coded such that '1' represents the worst and '6' the best grade

Table 2 Descriptive statistics for post-test scores on knowledge acquisition

Post-test scores	Points	PF	VF-process	VF-outcome	TOTAL
		<i>n</i> = 55 <i>M</i> (<i>SD</i>)	<i>n</i> = 56 <i>M</i> (<i>SD</i>)	<i>n</i> = 62 <i>M</i> (<i>SD</i>)	<i>N</i> = 173 <i>M</i> (<i>SD</i>)
Conceptual	0–7	2.24 (1.35)	3.03 (1.47)	2.52 (1.78)	2.59 (1.58)
Procedural	0–5	2.26 (2.01)	2.55 (1.98)	2.42 (1.92)	2.41 (1.96)
Overall	0–12	4.49 (2.77)	5.58 (3.07)	4.94 (3.21)	5.00 (3.05)

was actually *more* effective for conceptual knowledge acquisition than PF, $F(1, 166) = 6.91$, $p = 0.009$, $\eta_p^2 = 0.04$. However, we also found a significant interaction between the effect of condition and the students' prior knowledge, $F(2, 166) = 3.13$, $p = 0.047$, $\eta_p^2 = 0.04$, which indicates that the impact of the students' prior knowledge on their performance on the conceptual knowledge post-test differs between VF-process and PF. To follow up on this interaction effect, we conducted a regression model with the SPSS macro PROCESS (Model 1: see Hayes 2013). As can be seen in Fig. 3, prior knowledge significantly moderated the effect on conceptual knowledge in the PF and the VF-process conditions, $b_3 = 0.34$, $p = 0.011$ (for all regression results: see Table 3). While prior knowledge did not affect the post-test performance in the PF condition, $r(55) = 0.13$, $p = 0.345$, students in the VF-process condition gained more from observing the process when they had higher prior knowledge, $r(56) = 0.53$, $p < 0.001$. Using the Johnson-Neyman technique (see Hayes 2013), we further identified regions in the range of the continuous moderator (i.e., prior knowledge), in which the effect between VF-process and PF on the students' conceptual knowledge reached statistical significance (level of significance: $\alpha = 0.05$). Our results revealed that the moderation effect exceeds the level of significance only if students had achieved at least 4.52 points out of a maximum 10 points on the prior knowledge pre-test ($n = 65$, or 58.56% of the students in the PF and VF-process conditions).

Taken together, the results of the regression model revealed that from a score of 4.52 points on the prior knowledge pre-test, the difference between VF-process and PF on the conceptual knowledge post-test increased by 0.34 points for each further point on the prior knowledge pre-test (see Fig. 3). In other words, there was only a significant difference between the VF-process and PF conditions for students who had a certain amount of prior knowledge.

With regard to our second hypothesis that students in the PF and VF-process conditions would outperform students in the VF-outcome condition, contrary to our expectation, the contrast showed no significant differences, $F(1, 166) = 1.42$, $p = 0.235$, $\eta_p^2 = 0.01$.

Procedural knowledge

As both mathematical ability, $r(173) = 0.36$, $p < 0.001$, and prior knowledge, $r(173) = 0.27$, $p < 0.001$, correlated positively with procedural knowledge acquisition, these measures were used as covariates in the analysis. In line with our expectations, an ANCOVA on procedural knowledge showed that there were no significant differences among conditions with respect to procedural knowledge acquisition, $F(2, 168) = 0.15$, $p = 0.863$, $\eta_p^2 = 0.002$. The ANCOVA also revealed a significant effect of prior knowledge, $F(1, 168) = 6.97$, $p = 0.009$, $\eta_p^2 = 0.04$, and mathematical ability, $F(1, 168) = 16.86$, $p < 0.001$, $\eta_p^2 = 0.09$.

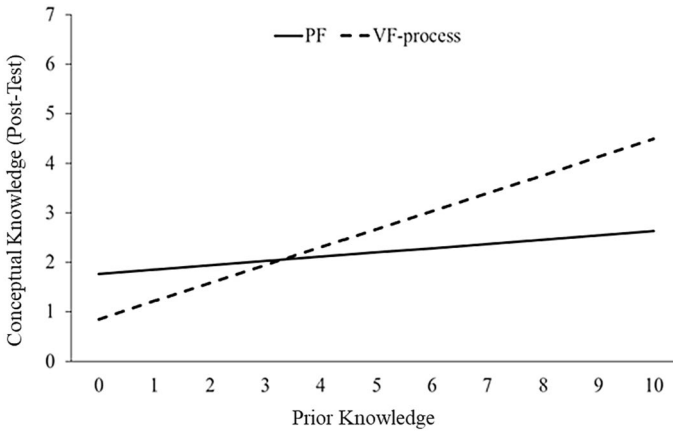


Fig. 3 Moderating effect of students' prior knowledge (0–10 points) on students' performance on the conceptual knowledge post-test (0–7 points)

Table 3 Regression results of the moderation analysis

Dependent variable: conceptual knowledge acquisition					
	<i>b</i>	<i>b</i> 95% CI [LL, UL]	<i>SE</i>	<i>t</i>	<i>p</i>
Constant (b_0)	2.30	[1.96, 2.64]	.17	13.28	< .001
Condition (b_1)	0.71	[0.22, 1.20]	.25	2.89	.005
VF-process (1)					
PF (0)					
Mathematical ability (Covariate)	0.36	[0.11, 0.62]	.13	2.79	.006
Prior knowledge (b_2)	0.04	[-0.13, 0.21]	.09	0.47	.639
Condition \times prior knowledge (b_3)	0.34	[0.08, 0.60]	.13	2.58	.011

Awareness of knowledge gaps

Because awareness of knowledge gaps was highlighted as an important preparatory mechanism underlying the effectiveness of the PF approach (Loibl and Rummel 2014a), we expected that the self-reported awareness of knowledge gaps would be positively associated with students' conceptual knowledge acquisition (Hypotheses 1 and 2), respectively, students in the PF and VF-process conditions would report more knowledge gap awareness than students in the VF-outcome condition. However, our data did not support this assumption. An ANOVA did reveal a significant main effect of condition on awareness of knowledge gaps, $F(2, 170) = 11.57$, $p < 0.001$, $\eta_p^2 = 0.12$; yet, Bonferroni post-hoc tests showed that PF students reported significantly more awareness of knowledge gaps ($M = 2.88$, $SD = 1.07$) than did students in both the VF-process ($M = 2.06$, $SD = 1.00$, $p < 0.001$) and the VF-outcome condition ($M = 2.15$, $SD = 0.91$, $p < 0.001$). Additionally, there was no significant correlation between students' awareness of knowledge gaps and their performance on the conceptual knowledge post-test, either overall $r(173) = -0.09$, $p = 0.223$ or separated

by condition (PF: $r(55) = -0.12$, $p = 0.375$; VF-process: $r(56) = -0.08$, $p = 0.538$; VF-outcome: $r(62) = -0.06$, $p = 0.618$).

Quantity and quality of solution attempts

PF students generated $M = 2.95$ ($SD = 1.25$) solutions on average, and their solutions included $M = 1.04$ ($SD = 0.67$) components of the canonical formula. Notably, these results differ from previous research on PF. For instance, Loibl and Rummel (2014b), who used the same materials, reported an average of 4.13 ($SD = 1.35$) generated solutions, including 2.13 ($SD = 0.83$) canonical components. Neither solution quality, $r(55) = 0.10$, $p = 0.453$, nor quantity, $r(55) = 0.12$, $p = 0.374$ were significantly correlated with conceptual knowledge acquisition.

Discussion

The aim of our experiment was to compare the effectiveness of students generating their own solutions prior to instruction (PF) to students observing others attempting to generate solutions (VF-process) and to students looking at others' failed solution attempts (VF-outcome). We hypothesized that: (1) observing the problem-solving process prior to instruction (VF-process) would be as effective in terms of preparing students for learning from the subsequent instruction as engaging in problem solving (PF), and (2) students in the PF and VF-process conditions would show more conceptual knowledge in the post-test than students who only looked at the outcome of the problem-solving process (VF-outcome).

With regard to our first hypothesis, we found that students in the VF-process condition even outperformed students in the PF condition on the conceptual knowledge post-test. We also found an interaction with prior knowledge, revealing that students in the VF-process condition gained more from observing the process when they had higher prior knowledge. Regarding our second hypothesis, we did not find that students in the PF and VF-process conditions significantly outperformed students in the VF-outcome condition.

Our findings replicate and extend the results of research on example-based learning in the PF paradigm. First and foremost, while past research seemed to indicate that VF was less effective than PF Kapur (2014a, b), our findings show that it can be as effective or even more effective when VF shows the entire process and not merely the outcome. Moreover, our findings add to the results of previous studies on example-based learning, which showed that studying erroneous examples may foster students' conceptual knowledge acquisition even better than students' own problem solving (cf. McLaren et al. 2015), and that examples can be effectively used prior to instruction (cf. Glogger-Frey et al. 2015; Likourezos and Kalyuga 2017). Furthermore, our finding that students' prior knowledge moderated the effect on conceptual knowledge acquisition in the PF and the VF-process conditions is in line with the results of Große and Renkl (2007), who also revealed that prior knowledge affects learning from erroneous examples. However, the examples used in the studies by Glogger-Frey et al. (2015) and Likourezos and Kalyuga (2017) allowed students to study how to apply essential parts of the canonical solution prior to instruction. In contrast, the examples used in our study showed a student who was attempting and failing to solve the problem. Our findings show that example-based learning can still be effective in that case, and indeed even more effective than attempting to solve the problem

oneself, particularly for students with higher prior knowledge, and provided that students get to observe the entire problem-solving process. The question therefore arises of which mechanisms explain the superior effectiveness of the VF-process compared to PF.

The core mechanisms for the effectiveness of the PF approach that have been identified are prior knowledge activation, awareness of knowledge gaps, and deep feature recognition (Loibl et al. 2017). Regarding prior knowledge activation, previous studies found a positive correlation between the number of solution attempts generated during PF and conceptual knowledge acquisition (e.g., Kapur 2012; Kapur and Bielaczyc 2012). However, the findings are mixed, as Loibl and Rummel (2014b) did not confirm this correlation. We expected that students would also effectively activate prior knowledge while observing another student engaged in the problem-solving process. The model student displayed several components of the canonical solution in the problem-solving attempts, which would allow for extensive prior knowledge activation in VF. In contrast, students in the PF condition in the present study did not seem to engage in extensive prior knowledge activation, and their prior knowledge activation did not correlate significantly with their performance on the conceptual knowledge post-test. PF students in our sample generated less solution attempts with less components of the canonical solution included, compared to a study by Loibl and Rummel (2014b) and Hartmann et al. (2020), who used the same materials.

The fact that the PF condition in our sample produced less solution attempts with lower quality was surprising, because PF was designed according principles formulated by Kapur and Bielaczyc (2012). Even though we have implemented PF similar to previous studies, however, in the most studies on the PF approach students were engaged in collaborative problem solving. Accordingly, in some studies, students were spatially separated to enable an effective small group collaboration (e.g., Loibil and Rummel 2014a, b). In our study, approximately 20 students of a certain class first were randomly assigned to the experimental conditions and then worked individually in one classroom. Although previous studies have shown beneficial effects of the PF approach even if students worked individually (Kapur 2014a) in a classroom setting, it might have been easier for students to not continue generating any more solution attempts if it became too demanding for them, without being noticed.

Nevertheless, assuming that the quantity and quality of solution attempts reflects the range (or diversity) of students' prior knowledge activation (Kapur 2012), and assuming that prior knowledge activation is important to benefit from subsequent instruction (Loibl et al. 2017), this may explain why students in the PF condition did not acquire as much conceptual knowledge as students in the VF-process condition. However – and this is a limitation of the present study – we had no direct measure of prior knowledge activation in all conditions, meaning that this explanation remains speculative.

We did measure students' awareness of knowledge gaps, which is considered necessary in order to learn more effectively from the subsequent instruction (Loibl and Rummel 2014a). However, these results do not help explain why students in the VF-process outperformed those in the PF condition, because students in the PF condition reported more awareness of knowledge gaps than did VF students, and additionally, we found no significant correlation between the students' self-reported awareness of knowledge gaps and their performance on the conceptual knowledge post-test. This corresponds with results by Glogger-Frey et al. (2015) showing that more knowledge gap awareness through problem-solving prior to instruction did not result in higher learning outcomes. Glogger-Frey et al. (2015) argued that perhaps, during instruction, students might not have been able to concentrate on the relevant information, which would have been essential to specifically address their knowledge gaps. This might also explain our findings. Another explanation

might lie in the fact that there is no linear relation between the students' awareness of knowledge gaps and how effectively they learned from instruction. Even though students in the PF condition were more aware of their knowledge gaps, the awareness as triggered by studying examples might have been sufficient to benefit from the subsequent instruction.

It should be noted, however, that the knowledge gap awareness items are rather unspecific, and they do not assess explicit knowledge about why a particular solution attempt did not adequately solve the problem (i.e., negative knowledge). While PF students might have identified that their solution attempts led to erroneous results, thus becoming more aware of their knowledge gaps, they may not have been able to explicitly identify why their attempted solutions were ineffective (or incomplete). This brings us to the third mechanism: deep feature recognition. An alternative approach to explaining why VF-process reported less awareness of knowledge gaps, yet benefited more from their prior knowledge and outperformed PF on the conceptual knowledge post-test, would be that observing another student's solution attempts sets free resources to reflect more effectively on the deep features of the solution attempts. In contrast to generating one's own solutions, studying another student's solutions requires observers to face information, which might be inconsistent with their own prior knowledge. This might potentially lead to cognitive conflict (Limón 2001), that is, identifying discrepancies between the model's intuitive ideas on how to solve the problem and the observer's naïve conceptions. Once the observer identifies, attempts to explain, and solves such conflict, the observer can use his or her prior knowledge to reflect on the rationale behind the model's actions, in order to understand why certain solutions do not work (i.e., negative knowledge). In this way, the observer may gain insight into the deep features of the concept to be learned. In other words, because VF-students did not have to generate their own solution attempts, they might have had more cognitive resources available for gaining explicit knowledge about why (some) of the components of the canonical solution are important, prior to the instruction. The fact that more components of the canonical solution (i.e., better quality) were displayed to VF-students could have increased this advantage, especially for students with higher prior knowledge, as they would have been able to identify better solutions even if they could not have generated them themselves. It has not yet been investigated, however, whether recognizing deep features of a not-yet-known concept already plays a role during problem solving or exclusively takes place during the subsequent instruction (Loibl et al. 2017). If students already recognize some of these deep features of the concept to be learned during PF or VF, it will be easier for them to understand and learn the canonical solution components as taught during the instruction. This would also explain why the VF-process was particularly effective for students with higher prior knowledge. Higher prior knowledge both reduces experienced cognitive load imposed by the task (Kalyuga 2011) and aids in the deep feature recognition: When observers have lower prior knowledge, they might not be able to identify and explain the discrepancies between their own and the observed concept (Limón and Carretero 1997). This explanation also corresponds to a study by Koedinger and Anderson (1990) showing that problem-solving experts (cf. students with higher prior knowledge in our study) focus more on key features of a problem than individuals with less expertise.

In contrast, PF students, who are generating their own solution attempts, can only gain insight into deep features during the problem-solving phase by effectively comparing and contrasting their own diverse solution attempts. This would not only require sufficient working memory capacity available to do so, but would also require there to be something to compare, and the low solution diversity observed in the present study suggests that this was not the case. This might also explain why some previous studies reported a positive

correlation between the students' generated solutions during the problem-solving phase and their performance on the conceptual knowledge post-test (Kapur 2014b; Kapur and Bielaczyc 2012), while some did not (Loibl and Rummel 2014b), because the effectiveness of the generated solutions might depend on further processes, like building negative knowledge and deep feature recognition. From this point of view, however, it could also be argued that the effectiveness of the PF approach would be explained not so much by an activation of prior knowledge as by a process of knowledge building. The effectiveness of being engaged in problem-solving attempts is, therefore, based less on an activation of intuitive ideas than on deriving deep features of the problem as well as problem solutions through an active exploration. Again, however, this explanation is rather speculative, and tapping into these processes more directly during PF and VF is an important avenue for future research.

Conclusion

In conclusion, our data support the assumption that observing someone else's problem solving can prepare students for the subsequent instruction more effectively compared to one's own problem solving. In contrast to the assumption that only by generating their own solutions can students explore a not-yet-known concept Kapur (2014a, b), our experiment revealed that observational (or example-based) learning also plays an essential role in explaining the beneficial effects of problem solving prior to instruction. However, further research is needed to examine how students utilize their prior knowledge to elaborate on solution attempts during generation or observation. To examine why observing examples of failure prepares students with higher prior knowledge even more effectively for instruction than engaging in one's own (failed) problem-solving attempts, it would be fruitful to understand why (observing or actually) 'failing' helps students to 'productively' gain conceptual knowledge from subsequent instruction. Accordingly, to explore underlying preparatory effects by the means of qualitative process analyses would be a fruitful methodological approach for future studies. As we found that generating one's own solutions was not essential for gaining more from the PF approach, our findings have an important implication for educational practice: Students might be supported in 'failing productively' by observing examples, especially if they would otherwise struggle to generate diverse solution attempts. However, in general, we need to come to understand the underlying preparatory mechanisms of the PF approach in more detail to be able to carefully design pedagogical practices, which effectively and sustainably prepare students for instruction.

Acknowledgements The authors would like to thank all schools and students for their participation. We would also like to thank our student assistant Charleen Brand for her outstanding commitment to data collection.

Funding Open Access funding enabled and organized by Projekt DEAL.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The data collection followed the standards of the German Psychological Society (DGPs). According to the guidelines, no ethical approval was necessary at the time of conducting the study (2017) given the type of data we collected.

Informed consent The authors obtained written parental consent for all participants.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

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>.
- Ashman, G., Kalyuga, S., & Sweller, J. (2019). Problem-solving or explicit instruction: Which should go first when element interactivity is high? *Educational Psychology Review*. <https://doi.org/10.1007/s10648-019-09500-5>.
- Brand, C., Hartmann, C., & Rummel, N. (2018, June 23–27). Exploring relevant problem-solving processes in learning from productive failure. In J. Kay, R. Luckin (Eds), *Rethinking learning in the digital age: Making the learning sciences count, 13th International Conference of the Learning Sciences (ICLS)* (pp. 1125–1128), Vol. 2. London, UK: International Society of the Learning Sciences. Retrieved December 14, 2020, from <https://repository.isls.org/handle/1/576>.
- 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>.
- 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. <https://doi.org/10.1016/j.learninstruc.2011.11.001>.
- 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>.
- 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. <https://doi.org/10.1016/j.learninstruc.2007.09.008>.
- Hartmann, C., Van Gog, T., & Rummel, N. (2020). Do examples of failure effectively prepare students for learning from subsequent instruction? *Applied Cognitive Psychology*, *34*, 879–889. <https://doi.org/10.1002/acp.3651>.
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: The Guilford Press. <https://doi.org/10.1111/jedm.12050>.
- 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., & 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>.
- Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*, 1–19. <https://doi.org/10.1007/s10648-010-9150-7>.
- Kalyuga, S., & Singh, A.-M. (2016). Rethinking the boundaries of cognitive load theory in complex learning. *Educational Psychology Review*, *28*, 831–852. <https://doi.org/10.1007/s10648-015-9352-0>.
- 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. *Journal of the Learning Sciences*, *23*, 651–677. <https://doi.org/10.1080/10508406.2013.819000>.

- Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, 21(1), 45–83. <https://doi.org/10.1080/10508406.2011.591717>.
- Koedinger, K. R., & Anderson, J. R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, 14(4), 511–550. https://doi.org/10.1207/s15516709cog1404_2.
- 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>.
- Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and instruction*, 11(4–5), 357–380. [https://doi.org/10.1016/S0959-4752\(00\)00037-2](https://doi.org/10.1016/S0959-4752(00)00037-2).
- Limón, M., & Carretero, M. (1997). Conceptual change and anomalous data: A case study in the domain of natural sciences. *European Journal of Psychology of Education*, 12(2), 213–230. <https://doi.org/10.1007/BF03173085>.
- 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., 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>.
- Lu, J., Bridges, S., & Hmelo-Silver, C. (2014). Problem-based learning. In R. Sawyer (Ed.), *The cambridge handbook of the learning sciences*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139519526.019>.
- 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. (2002). Learning from worked-out examples: Instructional explanations supplement self-explanations. *Learning and Instruction*, 12, 529–556. [https://doi.org/10.1016/S0959-4752\(01\)00030-5](https://doi.org/10.1016/S0959-4752(01)00030-5).
- Renkl, A., & Eitel, A. (2019). Self-explaining: Learning about principles and their application. In J. Dunlosky & K. A. Rawson (Eds.), *The cambridge handbook of cognition and education*. Cambridge: University Press. <https://doi.org/10.1017/9781108235631.022>.
- Seger, C. A. (1994). Implicit learning. *Psychological bulletin*, 115(2), 163–196. <https://doi.org/10.1037/0033-2909.115.2.163>.
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, 2(1), 59–89. https://doi.org/10.1207/s1532690xci0201_3.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>.
- 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., & 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*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108235631.009>.
- 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>.
- Wiedmann, M., Leach, R. C., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention? *Instructional Science*, 40, 711–730. <https://doi.org/10.1007/s11251-012-9204-y>