



How preparation-for-learning with a worked versus an open inventing problem affect subsequent learning processes in pre-service teachers

Inga Glogger-Frey^{1,2} · Anne-Katrin Treier^{1,3} · Alexander Renkl¹

Received: 16 February 2021 / Accepted: 29 January 2022 / Published online: 16 March 2022
© The Author(s) 2022

Abstract

A worked-out or an open inventing problem with contrasting cases can prepare learners for learning from subsequent instruction differently regarding motivation and cognition. In addition, such activities potentially initiate different learning processes during the subsequent (“future”) learning phase. In this experiment ($N=45$ pre-service teachers), we aimed to replicate effects of earlier studies on learning outcomes and, on this basis, to analyze respective learning processes during the future-learning phase via think-aloud protocols. The inventing group invented criteria to assess learning strategies in learning journals while the worked-example group studied the same problem in a solved version. Afterwards, the pre-service teachers thought aloud during learning in a computer-based learning environment. We did not find substantial motivational differences (interest, self-efficacy), but the worked-example group clearly outperformed their counterparts in transfer ($BF_{+0} > 313$). We found moderate evidence for the hypothesis that their learning processes during the subsequent learning phase was deepened: the example group showed more elaborative processes, more spontaneous application of the canonical, but also of sub-optimal solutions than the inventing group (BFs around 4), and it tended to focus more on the most relevant learning contents. Explorative analyses suggest that applying canonical solutions to examples is one of the processes explaining why working through the solution leads to higher transfer. In conclusion, a worked-out inventing problem seems to prepare future learning more effectively than an open inventing activity by deepening and focusing subsequent learning processes.

Keywords Invention as preparation for learning · Worked example · Teacher education · Assessment of learning strategies · Think-aloud · Learning processes

✉ Inga Glogger-Frey
inga.frey@uni-erfurt.de

¹ Department of Educational and Developmental Psychology, Institute of Psychology, University of Freiburg, Freiburg, Germany

² Department of Psychology, University of Erfurt, Erfurt, Germany

³ School for Child and Adolescent Cognitive Behavior Therapy (AKiP), Medical Faculty, University of Cologne, Cologne, Germany

Introduction

How would you, as a teacher or lecturer, set the stage before you introduce new concepts? What cognitive or motivational processes can be elicited by different preparation methods? The present study focusses on such processes activated by different forms of preparation tasks and on whether they might explain different learning outcomes: Inventing with contrasting cases versus a worked solution of the invention problem (in short: worked example).

Inventing is a brief, problem-oriented starting method that is commonly used as a learning activity preceding and thereby optimizing direct forms of instruction (Glogger-Frey et al., 2015a; Schmidt et al., 1989; Schwartz & Bransford, 1998). It usually consists of several (contrasting) cases taken from the learning domain, along with the task to invent a principle (e.g., index, formula, criteria), which encourages to compare and contrast the cases (e.g., Roll et al., 2011, 2012; Schwartz et al., 2011).

The present study's learners were pre-service teachers, and the domain is the assessment of learning strategies. This domain is important for (pre-service) teachers, who are expected to support students in their learning and in becoming life-long learners. Student cases representing their learning-strategy use can be compared and contrasted. The strategy cases encompass strategies of different qualities so that pre-service teachers can invent evaluation criteria by contrasting and comparing these cases.

This activity can provide motivational as well as cognitive preparation to learn from a future learning phase. However, the open task of *inventing* criteria could also have disadvantages when learners meander with sub-optimal solution attempts. Therefore, more guidance such as working through a worked-out solution of the same problem and the same material, that is, a *worked-example activity* has been proposed (Ashman et al., 2020; Glogger-Frey et al., 2015b; Hsu et al., 2015; Newman & DeCaro, 2019; Roelle & Berthold, 2016). Glogger-Frey, et al. (2015a), for example, found that it can be more beneficial regarding self-efficacy and transfer. This result was rather surprising and the question arises, how learning from the subsequent, "future" learning phase is altered by the different starting methods, that is, what the mechanisms of the effect are. In the present study, we aim to shed light on the learning processes during the future learning phase that are activated by the two preparation tasks, potentially explaining why learning outcomes differ. As a basis for this aim, we strive to replicate core findings of previous studies.

Motivational processes activated by preparatory activities

Previous studies have found that an *inventing activity* activated processes that affect motivational states, for example, situational interest (Glogger-Frey et al., 2015a, b; Rotgans & Schmidt, 2011, 2014; Weaver et al., 2018). Situational interest is a state that is externally triggered by features of the learning situation (Hidi & Berndorff, 1998). Situational interest is especially maintained if a topic or material is, for example, useful to the learner (value-related interest). An activity such as inventing can make pre-service teachers more aware of the *incompleteness* of their knowledge than working with a worked example (Glogger-Frey et al., 2017; Glogger-Frey et al., 2015a; Ribeiro et al., 2018): while contrasting the authentic student cases, pre-service teachers might realize that they do not know how learning strategies can be evaluated. They might realize that it would be useful to know for their teaching. Rotgans and Schmidt (2014) conducted a series of studies supporting the hypothesis that realizing knowledge gaps increases situational interest. For example, students who

showed awareness that they lacked knowledge to understand learning material reported higher situational interest. Thus, we hypothesize that realizing knowledge gaps and the usefulness of learning contents can increase situational interest.

A *worked example activity* can trigger motivational processes enhancing self-efficacy. *Self-efficacy* is defined as one's beliefs about one's own capability to perform a particular task (Bandura, 1997). Observing how a problem is solved in a worked example can make learners believe that they can likely solve the task themselves (Bandura, 1997; Schunk, 1995; Sitzmann & Yeo, 2013). Accordingly, van Harsel et al. (2019) found that when examples preceded open problem-solving in the beginning of a learning sequence, learners' self-efficacy was higher than when they started with problem-solving. In the present case, pre-service teachers' belief that they could assess strategy cases by relevant criteria can get stronger while working through cases that are already evaluated using such criteria. Thus, we can hypothesize that *self-efficacy* is higher when working with a worked solution (i.e., worked example) than with an open problem. Self-efficacy was found to predict learning outcomes (Bandura, 1997; Glogger-Frey et al., 2015a). Hence, if a worked example enhances self-efficacy, this in turn could contribute to knowledge acquisition and lead to enhanced learning outcomes.

Cognitive processes activated by preparatory activities

There is a controversy between proponents of discovery learning and protagonists of direct instruction concerning cognitive advantages of a brief, problem-oriented learning activity such as inventing as preparation to learn. Invention provides little assistance compared to a worked-out problem containing the solution; that is, basically, there is disagreement on how much assistance is needed (cf. the assistance dilemma, Koedinger & Aleven, 2007). Inventing proponents postulate that inventing helps learners with little prior knowledge in particular because it generates early forms of knowledge.¹ This early knowledge facilitates the assimilation of new information during future learning (Schwartz & Martin, 2004). Inventing proponents also postulate that if learners reach an impasse in trying to solve the invention problem, they can perceive specific "information needs" (Glogger-Frey et al., 2015a; Loibl et al., 2017). As a consequence, learners should *focus on relevant contents in a subsequent learning phase* (Glogger-Frey et al., 2015a). Research on impasse-driven learning has shown that instructional explanations are more effective when given in the context of such an impasse (Sánchez et al., 2009; VanLehn et al., 2003).

In contrast, protagonists of direct instruction argue that especially learners with low prior knowledge should be directly guided to the canonical solutions and key concepts, avoiding erroneous learning paths that cost valuable time and cognitive resources (Kirschner et al., 2006; Sweller et al., 2007). Learners do not always generate canonical solutions during inventing and they are not guided directly to key conceptual information of the learning domain. A worked-out solution of the invention problem (worked example) allows the learners to directly process key information and the canonical solution, thus building a first consistent knowledge base. In line with this argumentation, two experiments showed that inventing lead to lower transfer of the acquired skills than the worked example (Glogger-Frey et al., 2015a).

¹ A closely related discussion can be found in literature on *problem-solving before instruction* or *productive failure*, see Appendix A.

<p>Case 1</p> <p>If I think about it, I find the topic mitosis quite exciting 'cause without mitosis nothing could grow, not even me! Just because of tiny genes do all these particles (chromosomes and chromatids) go to where they should and everything is split into two cells. These must be crazily complicated processes.</p> <p>Notes:</p>	<p>I (Expert teachers) evaluate (1) the degree to which learning and understanding is fostered as:</p> <p><input checked="" type="checkbox"/> high <input type="checkbox"/> middle <input type="checkbox"/> low</p> <p>Rationale: self-made and detailed/elaborated</p>	<p>Case 2</p> <p>In class we learned that it is exciting to know how mitosis works because that is the prerequisite that's why everything grows or isn't able to grow. Because of tiny little genes particles (chromosomes and then chromatids) go where they should and everything is split into two cells. These are complicated processes.</p> <p>Notes:</p>	<p>I (Expert teachers) evaluate (1) the degree to which learning and understanding is fostered as:</p> <p><input type="checkbox"/> high <input checked="" type="checkbox"/> middle <input type="checkbox"/> low</p> <p>Rationale: not self-made, but detailed/elaborated</p>
<p>Case 3</p> <p>If I think about it, it becomes clear to me that the topic mitosis is indeed exciting. 'Cause now I know how cells divide, how things grow!!! The mitosis happens so that one, that is a cell, slowly becomes two.</p> <p>Notes:</p>	<p>I (Expert teachers) evaluate (1) the degree to which learning and understanding is fostered as:</p> <p><input type="checkbox"/> high <input checked="" type="checkbox"/> middle <input type="checkbox"/> low</p> <p>Rationale: not detailed/elaborated, but self-made</p>	<p>Case 4</p> <p>Today we've discussed in biology that mitosis is quite exciting because I then know how cells are split and how things grow. Mitosis happens so that one cell becomes two cells, we discussed.</p> <p>Notes:</p>	<p>I (Expert teachers) evaluate (1) the degree to which learning and understanding is fostered as:</p> <p><input type="checkbox"/> high <input type="checkbox"/> middle <input checked="" type="checkbox"/> low</p> <p>Rationale: neither self-made nor detailed/elaborated</p>

Fig. 1 Worksheet with the Contrasting Cases Showing the Experimental Variation. Shaded in grey / green are Words that were Added (or Replaced in Parentheses) for the Worked-Solution Version. The Student Cases were Hand-written in the German Original. (Color figure online)

All in all, different cognitive and motivational advantages and disadvantages could explain the learning outcomes of the different preparatory activities. To test these mechanisms, learning processes during the future learning phase should be assessed.

Differential processes during future learning

The worked example activity might allow for building a more consistent first knowledge base, for example about which criteria can be used to evaluate learning strategies. Some core contents (i.e., criteria) can be connected to concrete examples (see Fig. 1 for learning strategy cases). This base might alter learning during the future learning phase. More specifically, it might allow for applying superior *cognitive learning processes* such as organization and elaboration, as well as *metacognitive regulation*. A first knowledge base facilitates to organize learning contents, for example, to differentiate important from less important contents. It may facilitate elaborating during future learning in that the already encoded cases and core contents can more easily be related to examples and new concepts in the later learning phase. There might be more room for metacognitive regulation when less erroneous learning paths and less suboptimal, invented solutions are activated and load the limited working memory. (colour figure online)

There are several reasons to assume that the canonical solution (here: criteria for learning strategy assessment) will be applied more often if learners worked through a solution as preparation activity as compared to an inventing activity. First, the canonical solution is more likely encoded than after inventing solutions so that learners can simply apply it more often when confronted with new student cases in the future learning phase. Second, two effects could reduce learners' application of canonical solutions after inventing their *own* solutions: According to the IKEA effect, self-made products (Norton et al., 2012) are valued higher than given products. Thus, self-generated, suboptimal solutions during

the invention phase can be valued higher than the given canonical ones. According to the continued-influence effect (Johnson & Seifert, 1994) the assimilation of new learning contents can be hindered if learners integrate false information in their knowledge base in the very beginning of a learning phase—which could happen during inventing. The learners then continue to hold on to their originally integrated information and have difficulties to let go of this initial idea. Thus, after inventing, learners could stick to their own suboptimal solution instead of taking up the directly instructed canonical one. On this basis, we can assume that the worked-example activity lets learners use the canonical solution more often during future learning.

In addition, we can assume that learners *spent less time learning* with the learning material that most directly relates to the canonical solution after an inventing activity. Note that the canonical solution represents the learning contents most relevant to the learning goal—here, understanding and being able to apply quality criteria to assess learning strategies. Glogger-Frey et al. (2015a) found that an inventing group spent less learning time in a self-regulated future learning phase and focussed less on the most relevant learning contents, namely the unit on quality criteria, than a group that worked through the same, but worked problem. This might be a kind of effort-justification effect: learners attempt to reduce cognitive dissonance when encountering contradictions between own developed solutions and the canonical solution (cf. dissonance theory, Festinger, 1957) and thus spend less time learning with the learning contents most related to the canonical solution (here: quality criteria) than learners who worked with the canonical solution right away. All in all, we can assume that an open problem reduces the time learners spend working with relevant contents and reduces thinking (e.g., elaborating) about key contents compared to a worked example.

Hypotheses and research questions

In this experiment with pre-service teachers, we aimed to replicate and extend findings on effects of different preparation tasks (worked example or open inventing) on transfer as well as motivational and cognitive preparation processes (Glogger-Frey et al., 2015a). We aimed to investigate whether there are processes in the subsequent learning phase that explain potential differences in transfer between the two types of preparation-for-learning. We tested the following hypotheses:

Replication hypotheses

- (1) Inventing with contrasting cases enhances pre-service teachers' (1a) interest in learning contents (i.e., assessment of learning strategies), and (1b) working through a worked example to the invention problem enhances self-efficacy compared to the respective other group (*motivation hypotheses*).
- (2) Inventing (little assistance) leads to lower transfer of the acquired skills than the worked solution (i.e., worked example) (*learning outcome hypothesis*).

Process hypotheses

- (3) We expected that the worked-example activity, as compared to inventing, leads to superior learning processes during the “future” learning phase:
 - (3a) More cognitive learning processes (organization, elaboration)
 - (3b) More metacognitive regulation
 - (3c) More application of the canonical solution (the criteria) to provided examples (here: examples of learning strategies in learning-journal extracts)
 - (3d) Less application of suboptimal or false solutions to provided examples.

On the background of the effect of effort justification (Festinger, 1957) and previous results (Glogger-Frey et al., 2015a), we hypothesized that the inventing group

 - (3e) Spends less learning time
 - (3f) Focusses less (learning time) on the most relevant learning contents.

In addition, we tested whether differences in learning processes could provide explanations for effects on transfer (*Exploratory Mediation Hypotheses*).

Method

Participants, learning domain, and design

Forty-five pre-service teachers (30 female, 15 male; $M_{\text{age}} = 22.53$, $SD = 2.17$) participated in this study. Most of them were in the third semester of their studies (42%); the mean number of semesters was $M = 5.16$ ($SD = 2.57$). They were compensated with 15 Euros for participation (duration: about 109 min). Pre-service teachers had not yet learned about learning strategies and learning journals in their study program. The learning domain was the assessment of learning strategies in learning journals written by high school students. By writing learning journals, for example, as homework after biology or mathematics classes, high-school students are encouraged to apply learning strategies (Nückles et al., 2020). For example, they can develop their own thoughts based on the new learning contents (elaboration strategy). Ideally, they can do this in a detailed way and on their own (and not “copied”), for example, “I realized that nothing can grow without mitosis, not even myself! Because (...)” Such an elaboration can be evaluated as high in quality (Dunlosky et al., 2013; Glogger et al., 2012). The *ultimate learning goal* was to enable pre-service teachers to assess the quality of elaboration strategies in learning-journal extracts representing such strategies, as well as to achieve an understanding of quality criteria that enabled them to explain their assessments. Pre-service teachers were randomly assigned to two conditions: “inventing” ($n = 22$; 15 female, $M_{\text{age}} = 22.95$; $SD = 2.52$) and “worked example” ($n = 23$; 15 female; $M_{\text{age}} = 22.13$, $SD = 1.74$).

Material

The materials used were largely identical to those used in Glogger et al. (2015, first experiment). The experimental variation was completely identical.

Materials for the preparation phase—experimental variation

On the first page of the paper-and-pencil material, *all* participants received a very brief introduction (134 words) about learning journals and the quality of learning strategies in general. It stated, for example, that students usually differ in how well they qualitatively apply learning strategies, and that “quality” here refers to how conducive to learning the strategy was applied by a student. The last sentence read: “For the promotion of learning strategies, but especially for the learning success of the students, it is important that we as teachers can assess and give feedback on the *quality* of individual learning strategies”. The instruction to the subsequent activity read as follows: “On pages B and C, you will find four extracts from learning journals. Each extract shows a variation of the same elaboration strategy (developing own thoughts) in a way a student from a biology class (dealing with the topic mitosis) could have applied it.” The extracts were similar in length, but differed systematically in two quality criteria: (a) detailed elaboration vs. wordy, but shallow elaboration, (b) self-made vs. copied from the lesson (see Fig. 1). Hence, by comparing and contrasting the strategies, criteria could be identified to assess learning strategies. Participants in the *inventing group* invented criteria to evaluate the quality of learning strategies applied in learning journals. They were prompted to contrast the four extracts, to rate the quality of each extract on a 3-point scale (*low*, *medium*, or *high*), to make notes on discerning aspects, and to generalize from these aspects to generic evaluation criteria for the learning strategy elaboration. They had to write down their criteria in a box labelled “my criteria”. They were also instructed to check whether or not the final criteria really worked to discern all extracts (cf. Roll et al., 2012).

Participants in the *worked example group* neither had to rate the extracts nor to invent the evaluation criteria. Instead, they were asked to carefully study the same extracts including evaluations as worked out by a (fictitious) experienced teacher. That is, the blanks were filled: the canonical criteria were written in the “my criteria” box, the quality of the four extracts was rated, and short notes about discerning aspects were written down (see Fig. 1). In summary, the two groups had exactly the same work sheets with four extracts. The difference between conditions only referred to the following aspect: The inventing group had to generate the solution; the worked example group was given the solution. In other words, the inventing group had to generate the core principles by contrasting cases, whereas the solution group worked through the given principles and the contrasted cases without generating. The two criteria (principles) were explained in more detail in the subsequent instruction (a computer-based learning environment). Both groups were given the same amount of time (15 min) for their preparation activity (this duration was shown to be sufficient in pilot studies).

Pretest and demographic questionnaire

A web-based pretest assessed participants’ topic-specific prior knowledge. Participants received up to five points for the two open questions ($\alpha=0.67$, e.g., “Which learning strategies can students apply by writing a learning journal?”). Two independent raters scored 25% of the pretests ($ICC=0.91$; we used the two-way mixed effects model with absolute agreement and single rater measurement to calculate $ICCs$ in SPSS). A demographic questionnaire assessed sex, age, and number of semesters in teacher education.

Motivational questionnaire

A questionnaire assessed the participants' interest and self-efficacy right after the preparation activities, that is, before the learning phase ("pre") and directly after the learning phase ("post"). Situational interest was measured by eight items with a 6-point rating scale (6: *absolutely true*): feeling-related and value-related interest were assessed by four items, respectively (Schiefele, 1991; Schiefele & Krapp, 1996) (e.g., feeling-related / value-related: "Learning how to evaluate learning strategies is boring / useful."). Two additional items were constructed ("I would like to learn more about the assessment of learning strategies.", "I would like to receive an email with further information about the assessment of learning strategies."). They were included in the scale because inter-item-correlations were medium to high and reliability was clearly higher when including them (pre: Cronbach's $\alpha=0.75$; post: $\alpha=0.85$). Self-efficacy was measured with four items that were constructed following Bandura (2006), on a scale from 0% to 100% (pre: Cronbach's $\alpha=0.83$; post: $\alpha=0.79$; e.g., "I am confident I can differentiate well-applied from poorly-applied learning strategies.").

Computer-based learning environment and learning time

The future-learning phase took place in a computer-based learning environment (CBLE) described in Glogger et al., 2013. It explained several sub-categories of elaboration strategies, how they improve comprehension, and how they can be identified in learning journals in three units. A subsequent unit explained the two quality criteria of elaboration strategies targeted in the preparation activity. The CBLE's instructional design followed principles of example-based learning. For example, basics about the learning strategy were introduced followed by examples; a subsequent example had to be self-explained. Elaborated feedback was available to learners after completing interactive tasks (multiple-choice tasks, or drag-and-drop classification tasks). The *quality criteria* had to be understood and applied to solve transfer tasks in the posttest. The explanations in this unit were clearly more elaborate than in the worked-solution activity. Various examples of all sub-categories of elaboration strategies were used to illustrate different quality levels of strategies. The exact same extracts as in the preparation activity were not among these examples. Learners could navigate freely. The *focus on the most relevant learning contents* was operationalized as the time learners spent in the quality criteria unit (average duration in minutes $M=13$, $SD=4$). This duration as well as the total time spent in the environment were logged by the software (average total time in minutes: $M=38.6$, $SD=8.1$).

Coding of learning processes during the future-learning phase

In order to analyze learning processes, participants talked out loud what they were thinking during learning in the CBLE. We transcribed and coded the think aloud data during learning the quality criteria, as we were interested in preparation effects for learning these criteria. Five recordings could not be transcribed because the audio quality was too weak. Two independent raters coded 25% of the think-aloud data ($\kappa=0.89$). We let the grain size of learning processes determine the segment size because the segment size depended on the coded category (see Glogger et al., 2012; Renkl, 1997). For example, the expression

of negative mood “oh no” was usually much shorter than the application of a criteria to a learning journal extract. Hence, we segmented and coded the journals in one step.

In the development of the coding scheme, we deduced categories from theories on cognitive processing in multimedia learning (Fiorella & Mayer, 2015; Kiewra, 2004; Weinstein & Mayer, 1986) and according to our hypotheses. Some categories were specified bottom-up, based on the data in an iterative, qualitative process (see Chi, 1997). Each think-aloud segment was assigned to one of seven learning process categories: rehearsal, organization, elaboration, two metacognitive (monitoring, planning and regulating) processes, and the application of quality criteria (two processes: applying *canonical* solutions; applying *sub-optimal* solutions). An overview of categories, brief descriptions, and examples for the categories are provided in Table 1.

Rehearsal was seen as superficial learning process such as pure re-reading of learning materials or repeating what was just said (almost) word-by-word. All other relevant categories were regarded as deeper learning processes. Organization strategies referred to making connections between learning contents explicit, separating important from unimportant information, commenting on the content’s structure (structuring), and summarizing learning contents. Elaboration comprised deepening and working out learning content, for example, by drawing inferences or providing examples by which the contents are integrated into prior knowledge. Metacognitive processes included (a) statements of understanding or statements of problems in understanding (i.e., comprehension monitoring), and (b) planning and regulating the learning process by formulating questions to answer in the following or stating what to do next (sometimes along with a reason). Applying quality criteria referred to applying the canonical criteria (i.e., self-made vs. copied from the lesson; detailed elaboration vs. wordy, but shallow elaboration) or own suboptimal criteria when thinking about and evaluating an example in the learning environment.

Posttest: transfer

The posttest measured transfer and consisted of five tasks. Each task consisted of a brief extract from a learning journal, representing one sub-category of an elaboration strategy. This sub-strategy was labeled. All extracts were new, that is, they were taken neither from the preparation activity’s contrasting cases nor from the learning environment. The tasks required transfer as an application of quality criteria to new instances of students’ learning strategies. Two tasks asked pre-service teachers to give a reason for a provided evaluation of a learning-journal extract. Three tasks asked them to evaluate an extract themselves *and* to give a rationale for their own evaluation ($\alpha = 0.69$). Participants evaluated the extracts by rating the quality of the strategy as low, medium, or high. Answers were rated on a 6-point scale ranging from 1 (*no conceptual understanding*) to 6 (*very clear conceptual understanding*; SOLO taxonomy by Biggs & Collis, 1982). Two independent raters scored 25% of the posttest ($ICC = 0.93$).

Procedure

The pre-service teachers worked on the web-based pretest in the week before the experiment (at least four days) in order to avoid knowledge-activation effects. On the day of the experiment, in our lab, they first filled in the demographic questionnaire. Next, they worked individually on the task that prepared the following learning phase: The inventing group invented criteria to evaluate the quality of learning strategies applied in a learning journal;

Table 1 Category scheme

Categories	Description	Example segment(s)
Rehearsal	Repeating word by word	"Namely self-made examples, but hardly explained; or very detailed, but copied from the lesson."
Organisation	Summarizing and structuring, building connections within the learning contents	"That is very similar to what was said earlier, with the level of detail"
Elaboration	Deepening and working out learning content through illustrations and building connections to prior knowledge	"So here I can see the gradation from before."
Metacognitive monitoring	Comprehension monitoring	"If we got low quality, we do not see visualizations or real explanations anymore" [the learner drew an inference] "Hmh?" (uncertain) "I don't quite understand, I gotta say" "I am not sure about that now." "I've understood this, basically, so, what makes up the quality in learning journals"
Regulation	Planning and regulating the learning process, remedial strategies	"But I would like to look at the first one again because I find it a bit difficult." "But I will now compare with the others before I rate them."
Applying canonical solutions	Thinking about an example in terms of canonical criteria (that were taught in the learning environment)	"It is not explained in any way what albinism is, it lacks detail." (criterion: detailed vs. wordy, but shallow) "Own line of thought."
Applying suboptimal solutions	Thinking about an example in terms of criteria that were not taught in the learning environment, and not relevant to the quality of elaboration strategies	"Well, I think this is mainly taken from the lesson." (criterion: self-made vs. copied from the lesson) "It is well arranged."
Irrelevant	Navigation; Description of journal extracts; Other	"Does not sound so crazily interested" "Explains the issue factually correct" "Next", "I make the picture big now." "And, now, in this example there is a tree diagram" "Gosh, it's pouring."

the worked-example group was asked to carefully study a worked version of the very same problem for 15 min (constant time-on-task). Subsequently, questionnaires assessed the participants' situational interest and self-efficacy. The think aloud procedure was explained and then shortly practiced following the recommendation of Ericsson and Simon (1993; e.g., first exercise: How many windows does your parents' house have?). The participants then worked on the CBLE without time limits (on average for 38.6 min, $SD = 8.1$). During this learning phase, they spoke out loud what they were thinking. In case participants fell silent the experimenter said "keep talking", sitting outside the learners' field of vision (Fox et al., 2011). After the learning phase, interest and self-efficacy was reassessed. Finally, participants worked on the posttest.

Results

We used Bayesian Analyses. Bayes factors (BF) were estimated in JASP with the default prior (Cauchy scale 0.707). Following Jeffreys (1961), we interpreted a BF of 1 as no evidence, BF values in the range between 1 and $< = 3$ as anecdotal, BF values between 3 and $< = 10$ as moderate, between 10 and $< = 30$ as strong, and between 30 and $< = 100$ as very strong. The data give no evidence that experimental groups differed in any demographic variables (all $BF_{10} < 0.6$) or prior knowledge (inventing: $M = 3.02$, $SD = 1.47$; solution: $M = 2.70$, $SD = 1.15$; $BF_{10} = 0.391$, error % = 0.021).

We only found some small insignificant correlations between prior knowledge and process and outcome variables (see Table 4 in Appendix B for these and all other intercorrelations between prior knowledge, process variables, and transfer). We thus did not include prior knowledge in the analyses. We calculated Bayesian Independent Samples T-Tests using JASP (Version 0.14.1, JASP Team, 2020).

Replication hypotheses: motivation and learning outcomes (hypotheses 1a-b and 2)

Table 2 presents the means and standard deviations of motivational variables, learning process variables (processes during future learning measured by think aloud), learning time in the most relevant unit of the CBLE on quality criteria, and transfer for the two experimental groups, together with test statistics and effect sizes (Cohen's d). In contrast to the original study, the estimated Bayes factor (alternative/null) suggested no evidence for a beneficial effect of inventing on interest (directly after the preparation task as well as after learning in the learning environment $BF_{s_0} < 0.5$, see Table 2) or a beneficial effect of working through a worked example on self-efficacy *after* learning in the learning environment. We found anecdotal evidence of a beneficial effect of the worked-example activity on self-efficacy measured directly after the preparation task, $BF_{+0} = 1.374$.

Most importantly, we could replicate the effects of the previous study regarding learning outcomes: the Bayes factor indicates very strong evidence for the hypothesis that the worked example condition outperformed the inventing condition in transfer. Specifically, BF_{+0} means that the data are approximately 313 times more likely to occur under this alternative hypotheses H_+ than under H_0 (assuming no differences between the groups).

Table 2 Means (standard deviations in parentheses) of process variables (motivation, learning processes during future learning phase), transfer and learning time in the experimental groups

	Inventing <i>M (SD)</i>	Worked solution <i>M (SD)</i>	<i>d</i>	<i>BF</i>	error %
Interest Pre ^a	4.92 (0.58)	4.80 (0.91)	-.16	0.426	~0.003
Interest Post ^a	4.95 (0.84)	4.86 (0.98)	-.10	0.382	~0.008
Self-efficacy Pre ^a	48.30 (16.13)	55.50 (15.45)	.46	1.374	<0.001
Self-efficacy Post ^a	68.58 (17.43)	69.24 (12.73)	.04	0.328	~0.011
Rehearsal	2.20 (2.91)	3.35 (3.23)	.37	0.912	<0.001
Organisation	4.75 (4.25)	5.15 (5.07)	.09	0.378	~0.002
Elaboration	7.50 (4.35)	11.25 (6.44)	.68	3.608	<0.001
Metacognitive					
Monitoring	4.50 (2.52)	5.55 (5.13)	.26	0.618	<0.001
Regulation	0.75 (.79)	0.30 (.47)	-.70	0.112 ^b	<0.001
Applying canonical solutions	6.20 (4.93)	10.25 (6.35)	.71	4.238	<0.001
Applying suboptimal solutions	10.10 (7.68)	16.35 (10.62)	.68	0.114 ^c	<0.001
Transfer	2.54 (0.56)	3.48 (0.91)	1.24	313.490	<0.001
Learning time total	37.41 (7.53)	39.78 (8.70)	.29	0.699	<0.001
Learning time on criteria ^d	12.36 (3.82)	14.39 (3.96)	.52	1.860	<0.001

6-point scales: interest: from 1 (*not true at all*) to 6 (*absolutely true*); transfer test: from 1 (*no conceptual understanding*) to 6 (*very clear conceptual understanding*). Self-efficacy: 11-point-scale from 0 to 100%. For all Bayesian tests except for the one for interest and applying suboptimal solutions, the alternative hypothesis specifies that the worked example group (WE) was greater than the inventing group (Inv), that is, BF_{+0} is reported. Bold values point to evidences for differences between groups

^aInterest and Self-efficacy before (Pre) and after (Post) the learning phase

^bContrary to what we expected in Hypothesis 3b, Inv was greater WE. The BF for the data to occur under the hypothesis $Inv > WE$ is $BF_{-0} = 3.844$, error % < .001

^cContrary to what we expected in Hypothesis 3d, WE was greater Inv. The BF for the data to occur under the hypothesis $WE > Inv$ is $BF_{+0} = 3.458$, error % < .001

^dLearning time (in minutes) in the most relevant unit of the CBLE on quality criteria

Learning processes during the future learning phase and exploratory mediation analyses (hypotheses 3a–f)

The data provide moderate evidence for the elaboration part of Hypothesis 3a: During learning, the worked example group showed more cognitive processes of elaboration than the inventing group, $BF_{+0} = 3.608$ (see Table 2). We found no evidence for a difference between groups in rehearsal, organization, or monitoring processes. Metacognitive regulation was shown more often by the inventing condition than the worked-example group, contrary to what we expected (Hypothesis 3b). Since the Hypothesis 3b (worked example > inventing) is very unlikely to be true, we tested exploratively how probable the inverse hypothesis was. It was close to 4 times more likely that this data would occur under the hypothesis that the inventing group showed more metacognitive regulation than the worked example group (see Notes of Table 2).

During working with the various examples of learning strategies in the learning environment, the worked example group applied the canonical assessment criteria more often. The Bayes factor indicates moderate evidence for the corresponding Hypothesis 3c. In contrast

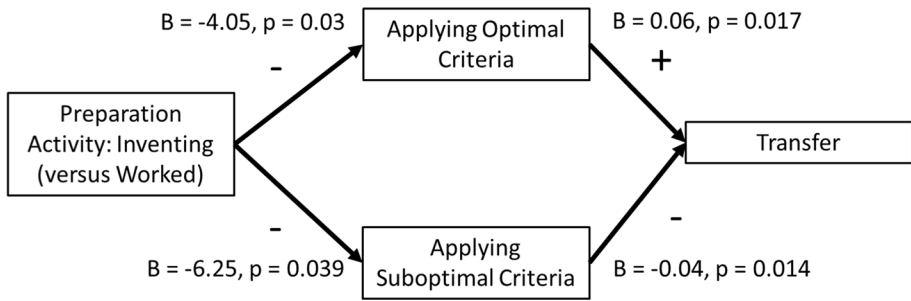


Fig. 2 Overview of the Mediation Effects: Learning Processes Significantly Explaining Higher Transfer Scores in the Worked Example Group (Exploratory Analyses; Worked Solution was Coded as zero, Inventing was Coded as one; Non-Standardized Coefficients "B" are Given for a- and b-Paths)

to Hypothesis 3d, the worked example condition applied suboptimal criteria to given examples *more often* than the inventing condition. We tested exploratively how probable the inverse hypothesis was. It was about 3.5 times more likely that the data occur under the hypothesis that the worked example activity leads to more application of suboptimal solutions than the inventing. The groups spend an equal amount of time with learning in the learning environment overall ($M = 38.60$ min, Hypothesis 3e rejected; the null hypothesis is 2.308 times more likely than the alternative hypothesis stating that the groups are different). We found anecdotal evidence for the hypothesis that the worked example group spent more learning time in the quality unit. That is, comparable with the previous study, the worked example group tended to focus more on the quality unit, that is, the most relevant learning contents, than the inventing group (Hypothesis 3f).

In exploratory analyses, we tested a multiple mediation model in SPSS using the PROCESS macro (Hayes, 2013). We aimed to test whether learning processes could explain the effect of conditions on transfer. We tested a model including the set of all variables for which we found (at least) moderate evidence for differences between the groups (these were also significantly different between conditions according to frequentist tests: elaboration, metacognitive regulation, application of canonical as well as suboptimal criteria). As in the standard settings of the Hayes (2013) process syntax, the mediation analysis generated 95% bias-corrected and accelerated bootstrap confidence intervals for the indirect effects using 5000 bootstrap samples.

If, as expected, the different preparation activities caused a change in learning processes that in turn caused changes in transfer, the indirect (mediation) effects via the learning-process variables should have confidence intervals not containing zero (Preacher & Hayes, 2008). We report partially standardized coefficients ("b") for the indirect effects.

Elaboration processes (a*b path: $b = -0.152$, 95% CI [-0.435, 0.024]) and metacognitive regulation (a*b path: $b = -0.114$, 95% CI [-0.103, 0.015]) did not mediate the effect of preparation activities significantly (model 4 of the process tool with four mediators). We found a significant indirect effect of the application of canonical criteria (a*b path: $b = -0.276$, 95% CI [-0.687, -0.012]) as well as of the application of suboptimal criteria (a*b path: $b = 0.292$, 95% CI [0.029, 0.761]). Figure 2 illustrates the mediation effects: Participants in the worked solution condition applied the canonical criteria more often, which in turn led to higher transfer scores. Participants in the inventing condition applied suboptimal criteria less often, which in turn led to higher transfer scores.

Discussion

This study aimed to replicate and more deeply explain findings of a previous study regarding effects of an open-inventing versus a worked-out-inventing (worked example) preparation activity by measuring and analyzing learning processes during the future learning phase. In other words, can we identify processes in the learning phase *following* the preparation that explain effects on transfer?

Most importantly, we found very strong evidence for the replicated transfer effect that learners who work through a canonical solution as preparation for learning achieve higher transfer scores than learners who invent own solutions to the problem. This finding is in line with previous findings and with the assumptions of protagonists favoring direct instruction (Hsu et al., 2015; Ashman et al., 2020; Glogger-Frey et al., 2015a; Newman & DeCaro, 2019; cf. Sweller et al., 2007). This finding formed the basis for our process analyses. We did not find strong motivational differences, but we found moderate evidence for the hypothesis that learning processes in the future-learning phase differed depending on how learners prepared for learning. These differences partly explain the effect on transfer.

Motivational processes

In contrast to the to-be-replicated study (Glogger-Frey et al., 2015a), we could not detect situational interest to be triggered more by the open inventing problem than by the worked problem. We expected this benefit of inventing because pre-service teachers realize what they do not know yet and they get aware of how the following learning material could be useful to close these gaps (Rotgans & Schmidt, 2014). Some studies have shown enhanced interest after working with an open inventing problem (Glogger-Frey et al., 2015a; inventing compared to reading solutions of peers: Glogger-Frey et al., 2015b). However, the number of suboptimal invented solutions was related to lower interest in the present (see Table A1) as well as in the replicated studies (where overall, nonetheless, interest was higher in the inventing group than in the worked group). Thus, there seem to be processes initiated by the inventing activity that dampen *and* elevate situational interest in comparison to a worked example. A worked example might also trigger interest: it provides basic key knowledge by illustrating the criteria in contrasting cases. Sufficient knowledge to organize new information is seen as a prerequisite for interest (Rotgans & Schmidt, 2017; Renninger, 2000; see Endres et al., 2020, for a similar argumentation). Thus, both variants of a preparation method have potential to trigger situational interest.

We found only anecdotal evidence for the hypothesis that self-efficacy was elevated by the model of a worked-out problem in comparison to open problem solving (inventing). The success rate of invention can provide an explanation for this. In the present study, 90% of pre-service teachers were at least partly successful (even though they often invented suboptimal criteria along with canonical ones, see reported results in Appendix A). Thus, they had good reasons to feel confident about assessing future learning-strategy cases (almost as confident as the worked example group). In Glogger-Frey et al. (2017), 56% of students found the canonical solution to a second inventing problem. No differences in self-efficacy between groups were found. In the second experiment in Glogger-Frey et al. (2015a), students' success rate was 28%; the worked-solution group showed higher self-efficacy which in turn mediated higher transfer. Thus, the success rate in inventing could be the crucial

factor in whether learners preparing with an open inventing problem feel comparable or less self-efficient than learners preparing with a worked example.

(Meta-)Cognitive learning processes in future learning: effects of worked examples

This study provides moderate evidence that working through the worked example deepens subsequent learning processes and focuses on the most important learning contents: learners in this condition elaborated more than their counterparts on the learning strategies and the criteria to evaluate them. They applied the canonical solution more often to strategy examples encountered during learning—but also sub-optimal solutions. According to the exploratory mediation analyses, applying canonical solutions to examples is one of the processes explaining why working through the solution leads to higher transfer.

Altogether and on the background of previous studies, we claim to see a direction even if future research definitely needs to provide more evidence: to prepare learning with worked examples seems to allow for deeper subsequent processes than the open inventing task. In addition, these deepened processes are focused on the most important learning contents (criteria to evaluate learning strategies). We will discuss this claim with regard to learning time as well as thinking processes.

Learning time

While the overall learning time did not differ between groups (37 versus 39 min on average), the worked example group spent a bit more time in the sections explicitly dealing with the criteria ($d=0.52$; anecdotal evidence). This effect was less pronounced than in the previous, replicated study (Glogger-Frey et al., 2015a: $d=0.85^2$). The effect might have been diluted in the present study because thinking aloud slows down thinking processes (Fox et al., 2011): all learners in the present study spent more time on learning overall (13 min) than in the replicated study (9 min) and the variances were higher (SD : 3.9 vs. 2.5).

Thinking processes

Learners in the worked example group focused learning processes more strongly on the most important learning contents not only by spending slightly more time with them. They also applied the crucial content, namely the criteria to evaluate learning strategies more often to the material presented in the CBLE. Thus, a tentative claim to be further investigated in future research is that preparing learning with a worked example can foster focused processing (Renkl 2015) in future learning. This finding is in line with the theoretical notion that learners should be directly guided to canonical, key concepts saving valuable time and cognitive resources compared to less guided forms of learning (Kirschner et al., 2006; Sweller et al., 2007).

² Standardized beta of $b^*=-0.39$ converted by an excel-file from <http://www.stat-help.com/spreadsheets/Converting%20effect%20sizes%202012-06-19.xls>, based on Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale, NJ: Erlbaum. pp. 281, 284, 285 and Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis*. New York, NY: Sage. pp. 239.

The result that the worked example group instead of the inventing group operates more often with the suboptimal solutions was unexpected against the background of the continued-influence effect and the IKEA effect. However, we might have to understand the application of suboptimal criteria—just like applying canonical criteria—as a kind of *elaboration*. Pre-service teachers applied criteria to presented material in the CBLE, that is, to learning-strategy examples. These criteria were just learned or part of student teachers' prior knowledge. That is, these processes categorized as “applying criteria” (both canonical and sub-optimal) can also be seen as elaboration of examples given in the learning environment. Other elaborative processes, categorized into the elaboration category, were also shown more often by the worked-example group. From this perspective, it is comprehensible that the worked example group also applied suboptimal criteria more often—dampening their learning to some extent (cf. exploratory mediation analyses).

The negative effect on transfer by applying non-canonical criteria is plausible. It can be further interpreted by conceptual change theories (diSessa, 2013; Vosniadou, 2013). These theories assume certain kinds of (suboptimal) prior knowledge that are difficult to address in learning. Pre-service teachers often show such knowledge (about learning strategies: Glogger-Frey et al., 2018; Ohst et al., 2014; about learning: Vosniadou et al., 2020; Eitel et al., 2021). In line with these theoretical approaches, holding on to such kind of knowledge impeded our participants' learning to some extent.

Hampering processes of inventing

Learning time and elaboration

Looking at the discussed findings from the angle of inventing, working through the open problem made pre-service teachers tend to spent less time with the most crucial learning contents and let them elaborate less (including application of criteria to examples, moderate evidence). This finding can be seen in line with the dissonance theory (Festinger, 1957). The inventing group might have attempted to reduce cognitive dissonance when encountering contradictions between own developed solutions and the solution presented in the CBLE. It thus invested less time and cognitive effort (i.e., elaboration) when engaging with the learning contents than the worked-example group.

Metacognitive regulation

Interestingly, metacognitive regulation was shown more often by the inventing condition (moderate evidence) and did not relate to higher transfer. As we show in the following, these thinking processes are more accurately understood as *regulation attempts* instead of *successful* regulation. We revisited the thinking-aloud segments coded as metacognitive regulation. It stood out that the segments often convey postponing a learning activity and going on in the learning content.³ So, it seems like learners attempt to regulate, but do not

³ Example segments are the following; only the segments in italic are the ones coded as metacognitive regulation, the rest is context for better comprehensibility: Example segment 1: “*But I’ll just put it back. Let’s have a look at the others*”. [the learner refers to learning-strategy examples]. Example segment 2: “So I assign it to the high quality for now. *But think about it again.*” Example segment 3: “Well, I find it hard to evaluate [...]. The middle one [the learner refers to medium quality], *so, well, I’ll have a look at the criteria again.*”.

often succeed, providing a potential reason for the lack of effect of these kind of learning processes on transfer.

Limitations and potential boundary conditions

The replicated transfer effect (and the following effects on processes) could have boundary conditions: it could be specific to the complexity of learning content for learners with specific levels of expertise. The learning domain in the present study is rather ill-structured—it is not trivial for pre-service teachers to identify that their solutions are not as appropriate as the scientifically proposed (canonical) ones. Identifying whether a well-structured mathematical task is correct or not is more straightforward. Particularly, it is not trivial for pre-service teachers because learners often hold fragmented, partly misconceptual knowledge about learning strategies as already mentioned (Glogger-Frey et al., 2018; Ohst et al., 2014). Roelle and Berthold (2016) found comparable effects as the present ones using contrasting cases holding a very high number of possible comparisons. These effects can be interpreted in line with Ashman et al. (2020): for learning with complex learning contents (or high element interactivity), direct instruction should precede problem-solving.

High element interactivity can also be due to low expertise (Chen et al., 2017). The present results might, thus, be specific for a certain degree of learner expertise combined with a certain element interactivity (complexity of learning contents). A comparable effect on transfer as in this study (i.e., worked example > inventing) was found in a rather well-structured domain, thus, with lower element-interactivity compared to the present contents (ratios in physics, e.g., density), but with less experienced learners, namely school students (Glogger-Frey et al., 2015a). The *opposite* effect (i.e., inventing > worked example) was found when such school students gained more expertise: school students prepared with two inventing problems (same topic); thus, they were given opportunity to practice. This practice made them having a slightly higher degree of expertise (Glogger-Frey et al., 2017), explaining the reverse effect (cf. the success rates mentioned above). Future studies should explicitly test this interpretation.

Apart from the very strong evidence for the effect on transfer, we did find moderate or lower evidences for our hypotheses. Thus, further experimental studies with greater sample sizes are necessary. Naturally, data collection with think aloud is very resource-intensive (individual sessions, transcription and coding of qualitative data). However, the current study proposes to look at specific processes, which could at least simplify the coding.

The current study looked at whether frequencies of learning processes differed between groups. There are good arguments to assume that the quality of learning processes would matter, too (Glogger et al., 2012; Roelle et al., 2017). For example, recognizing to a higher degree the overarching structure of the learning content could matter for learning outcomes. We actually made attempts in the coding phase of the study to measure the quality of processes. However, the think-aloud data and/or the measurement instrument revealed very restricted variances so we refrained from further analyses. Nevertheless, it could be fruitful for future research to take the quality of learning processes into account.

We argue that the study adds evidence for the more general assumption that in comparable learning domains, working through a worked example deepens future learning and thus is more beneficial for transfer than preparation with an open inventing problem. However, of course, such generalizations can be limited. Our sample are university students in a teacher education program, that is, usually interested in learning the target learning content

of this study (evaluating learning strategies). Whether we can generalize the effects regarding the level of expertise of learners and the learning domain has been discussed above.

We have used an immediate posttest. Thus, it is open whether the effects last over time. A delayed posttest could reveal different results.

Conclusion

In sum, we can argue that a worked example can foster learning processes such as elaboration and can focus the learners on the most important learning contents during a future-learning phase. This could provide a reason why, in a domain with rather high element-interactivity, we replicated the following effect with strong evidence: pre-service teachers preparing to learn with a worked-example activity clearly outperform their counterparts preparing with an open invention activity.

Appendix A

In order to provide analyses that are comparable with a related corpus of research, namely research on *productive failure* (e.g., Kapur, 2014, 2016) or *problem solving before instruction* (Loibl et al., 2017), we asked whether failure in the inventing activity would be productive (*Productive Failure Hypotheses*): We hypothesized that (4a) the appropriateness, (4b) the number of canonical, (4c) the number of suboptimal solutions, and (4d) the total number of invented solutions are related to learning outcomes.

Method: coding of inventing variables

In order to answer the research questions about the productiveness of suboptimal inventing (failure) or of the number of generated solutions, we rated the quality (appropriateness) of the invented solutions (i.e., in the inventing group only) on a 6-point scale ranging from 1 (*very low quality*) to 6 (*very high quality*; see Glogger-Frey et al., 2015a, b). The two independent raters scored 25% of the invented solutions ($ICC = 0.70$). We also counted solutions, operationalized by the number of (different) criteria invented by participants; we differentiated between (close-to) canonical and suboptimal solutions.

Results: correlation of inventing variables and learning outcomes (productive failure hypotheses 4a–d)

None of the inventing variables correlated substantially with transfer (Research questions 4a to 4d, r_s between -0.16 and 0.03 , all $p_s > 0.49$; controlling for prior knowledge, see Table 3). Twelve of 22 learners preparing with the open inventing problem invented one canonical solution. Eight of them invented even two canonical solutions. That is, 90% of learners seemed at least partly successful—however, all learners also invented suboptimal solutions, 50% invented 3 to 5 suboptimal solutions.

Exploring Table 3, we can see a significant negative correlation between inventing suboptimal solutions and interest before the learning phase ($r = -0.46$, $p = 0.035$) as well as the total time invested in learning ($r = -0.46$, $p = 0.034$). Inventing canonical solutions,

Table 3 Partial correlations in the inventing group between inventing variables and motivational variables, transfer, and learning-time variables, controlled for prior knowledge

	Appropriateness of invented solutions	Number of invented solutions (total)	Number of sub-optimal, invented solutions	Number of (close-to) canonical solutions
Interest Pre	.29	-.37	-.46*	.05
Interest Post	.15	-.27	-.26	-.09
Self-efficacy Pre	.34	.32	.12	.38 ⁺
Self-efficacy Post	.26	.24	.14	.22
Transfer	-.16	.001	.03	-.04
Learning time total	.27	-.38	-.46*	-.04
Learning time on quality criteria ^a	.02	-.22	-.18	-.12

$n=22$, $df=19$. Intercorrelations between think aloud measures rehearsal, organization, elaboration, meta-cognitive processes and inventing variables all $r < \pm .28$, with all $ps > .24$, with the exceptions elaboration and number of (close-to) canonical solutions: $r = -.41$, $p = .076$

^a Learning time (in minutes) in the most relevant unit of the CBLE on quality criteria

⁺ $p < .10$, * $p < .05$

however, had motivational and cognitive advantages: it was positively correlated with self-efficacy before the learning phase (marginally significant, $r=0.38$, $p=0.090$) and with elaboration processes during further learning (marginally significant, $r=-0.41$, $p=0.076$, see Notes of Table 3). These exploratory results suggest that in our contexts "failure" in the inventing phase had unproductive effects (cf. productive failure effects, e.g., Kapur, 2014).

Discussion productive failure hypotheses 4ad

The present study does not reveal findings that suggest productive failure. The number and appropriateness of invented solutions did not correlate with transfer. Other studies found that the number of solutions correlates positively with learning outcomes, and the quality (i.e., appropriateness) correlates negatively with learning outcomes (Hartmann et al., 2020, for example, find medium correlation coefficients). Our exploratory results suggest unproductive effects of inventing suboptimal solutions: pre-service teachers spent less time learning in the CBLE and were less interested the more suboptimal solutions they had invented. It must be noted, however, that the present instructional design differs from the "ideal" design of productive failure: for example, productive failure should involve a teacher-led consolidation phase in which the teacher guides students to contrast, assemble, and convert the relevant student-generated solutions into canonical solutions (Kapur, 2016). In addition, the domain (learning to assess learning strategies) might not be suitable for productive failure (cf. Loibl et al., 2017).

Appendix B

See Table 4.

Table 4 Intercorrelations of prior knowledge, process variables, and transfer

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Prior knowledge	-	.02	.02	-.11	-.25 ⁺	-.07	.05	.18	.02	.11	.27 ⁺	-.05	-.10
2 Interest Pre ^a	-	-	.74 ^{**}	.22	.24	.25	.10	-.15	-.25	.36 [*]	-.16	-.12	.08
3 Interest Post ^a	-	-	-	.14	.31 [*]	.22	.15	-.20	-.17	.19	-.10	-.14	.15
4 Self-efficacy Pre ^a	-	-	-	-	.62 ^{**}	.03	.12	-.10	.03	.06	.22	.08	.18
5 Self-efficacy Post ^a	-	-	-	-	-	.27 ⁺	.00	-.20	-.17	-.02	.03	-.23	-.03
6 Rehearsal	-	-	-	-	-	-	.40 [*]	.25 [*]	.02	-.13	.20	.20	.47 ^{**}
7 Organisation	-	-	-	-	-	-	-	.32 [*]	.24	-.20	.16	.24	.45 ^{**}
8 Elaboration	-	-	-	-	-	-	-	-	.59 ^{**}	-.10	.30 ⁺	.44 ^{**}	.53 ^{**}
Metacognitive	-	-	-	-	-	-	-	-	-	.15	.05	.41 ^{**}	.47 ^{**}
9 Monitoring	-	-	-	-	-	-	-	-	-	-	-.32 [*]	.16	.04
10 Regulation	-	-	-	-	-	-	-	-	-	-	-	.02	.11
11 Transfer	-	-	-	-	-	-	-	-	-	-	-	-	.69 ^{**}
12 Learning time total	-	-	-	-	-	-	-	-	-	-	-	-	-
13 Learning time on criteria ^b	-	-	-	-	-	-	-	-	-	-	-	-	-

N = 45

^aInterest and self-efficacy before (Pre) and after (Post) the learning phase^bLearning time (in minutes) in the most relevant CBLE unit on quality criteria⁺ $p < .10$. ^{*} $p < .05$. ^{**} $p < .01$

Acknowledgements We thank especially Christoph Ries, but also Nicole Dillner and Nadja for their contributions to data collection, and Rolf Schwonke for supervising students.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analyses were performed by the second and supervised by the first author. The manuscript was written by the first author, and all authors, especially the third author commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. No funding was received to assist with the preparation of this manuscript and for conducting this study.

Availability of data and material The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethical approval All participants gave informed consent prior to the experiment.

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

- Ashman, G., Kalyuga, S., & Sweller, J. (2020). Problem-solving or explicit instruction: Which should go first when element interactivity is high? *Educational Psychology Review*, 32(1), 229–247.
- Bandura, A. (1997). *Self-efficacy. The exercise of control*. W.H. Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. C. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Information Age Publishing.
- Chen, O., Kalyuga, S., & Sweller, J. (2017). The expertise reversal effect is a variant of the more general element interactivity effect. *Educational Psychology Review*, 29(2), 393–405.
- Chi, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of Learning Sciences*, 6(3), 271–315.
- diSessa, A.A. (2013). A bird's-eye view of the “pieces” vs “coherence” controversy (from the “pieces” side of the fence”). In S. Vosniadou (Ed.), *International Handbook of Research on Conceptual Change*. Educational psychology handbook series (pp. 31–48). New York: Routledge.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58.
- Eitel, A., Prinz, A., Kollmer, J., Niessen, L., Russow, J., Ludäscher, M., et al. (2021). The misconceptions about multimedia learning questionnaire: An empirical evaluation study with teachers and student teachers. *Psychology Learning & Teaching*, 147572572110287.
- Endres, T., Kranzendorf, L., Schneider, V., & Renkl, A. (2020). It matters how to recall—Task differences in retrieval practice. *Instructional Science*, 48(6), 699–728.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis. Verbal reports as data*. MIT Press.

- Festinger, L. (1957). *A theory of cognitive dissonance* (2nd ed.). Stanford University Press.
- Fiorella, L., & Mayer, R. E. (2015). *Learning as a generative activity. Eight learning strategies that promote understanding*. Cambridge University Press.
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*(2), 316–344.
- Glogger, I., Holzäpfel, L., Kappich, J., Schwonke, R., Nückles, M., & Renkl, A. (2013). Development and evaluation of a computer-based learning environment for teachers: Assessment of learning strategies in learning journals. *Education Research International*, *2013*(1), 1–12.
- Glogger, I., Schwonke, R., Holzäpfel, L., Nückles, M., & Renkl, A. (2012). Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies. *Journal of Educational Psychology*, *104*(2), 452–468.
- Glogger-Frey, I., Ampatziadis, Y., Ohst, A., & Renkl, A. (2018). Future teachers' knowledge about learning strategies: Misconcepts and knowledge-in-pieces. *Thinking Skills and Creativity*, *28*, 41–55.
- Glogger-Frey, I., Gaus, K., & Renkl, A. (2017). Learning from direct instruction: Best prepared by several self-regulated or guided invention activities? *Learning and Instruction*, *51*, 26–35.
- Glogger-Frey, I., Fleischer, C., Grüny, L., Kappich, J., & Renkl, A. (2015a). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction*, *39*, 72–87.
- Glogger-Frey, I., Kappich, J., Schwonke, R., Holzäpfel, L., Nückles, M., & Renkl, A. (2015b). Inventing motivates and prepares student teachers for computer-based learning. *Journal of Computer Assisted Learning*, *31*(6), 546–561.
- Hartmann, C., Gog, T., & Rummel, N. (2020). Do examples of failure effectively prepare students for learning from subsequent instruction? *Applied Cognitive Psychology*, *34*(4), 879–889.
- Hayes, A.F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Methodology in the social sciences. New York, NY, US.
- Hidi, S. & Berndorff, D. (1998). Situational interest and learning. In L. Hoffman, A. Krapp, K. Renninger, & J. Baumert (Eds.), *Interest and learning* (pp. 74–90).
- Hsu, C., Kalyuga, S., & Sweller, J. (2015). When should guidance be presented in physics instruction? *Archives of Scientific Psychology*, *3*(1), 37–53.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
- Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1420–1436.
- Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, *38*(5), 1008–1022.
- Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist*, *51*(2), 289–299.
- Kiewra, K. A. (2004). *Learn how to study and SOAR to success*. Pearson Prentice Hall.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86.
- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with Cognitive Tutors. *Educational Psychology Review*, *19*, 239–264.
- Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instructional supports learning. *Educational Psychology Review*, *29*(4), 693–715.
- Newman, P. M., & DeCaro, M. S. (2019). Learning by exploring: How much guidance is optimal? *Learning and Instruction*, *62*, 49–63.
- Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, *22*(3), 453–460.
- Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review*, *32*(4), 1089–1126.
- Ohst, A., Fondu, B. M. E., Glogger, I., Nückles, M., & Renkl, A. (2014). Preparing learners with partly incorrect intuitive prior knowledge for learning. *Frontiers in Psychology*, *5*, 664.
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879–891.
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*(1), 1–29.
- Renninger, K. A. (2000). *Individual interest and its implications for understanding intrinsic motivation* (pp. 373–404). Academic Press.

- Ribeiro, L. M. C., Mamede, S., Moura, A. S., de Brito, E. M., de Faria, R. M. D., & Schmidt, H. G. (2018). Effect of reflection on medical students' situational interest: An experimental study. *Medical Education*, 52(5), 488–496.
- Roelle, J., Nowitzki, C. & Berthold, K. (2017). Do cognitive and metacognitive processes set the stage for each other? *Learning and Instruction*, 54–64.
- Roelle, J., & Berthold, K. (2016). Effects of comparing contrasting cases and inventing on learning from subsequent instructional explanations. *Instructional Science*, 44(2), 147–176.
- Roll, I., Aleven, V. & Koedinger, K.R. (2011). Outcomes and mechanisms of transfer in invention activities. In L. Carlson, C. Hoelscher, & T.F. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 2824–2829). Austin, TX: Cognitive Science Society.
- Roll, I., Holmes, N. G., Day, J., & Bonn, D. (2012). Evaluating metacognitive scaffolding in guided invention activities. *Instructional Science*, 40(4), 691–710.
- Rotgans, J. I., & Schmidt, H. G. (2011). Situational interest and academic achievement in the active-learning classroom. *Learning and Instruction*, 21(1), 58–67.
- Rotgans, J. I., & Schmidt, H. G. (2014). Situational interest and learning: Thirst for knowledge. *Learning and Instruction*, 32, 37–50.
- Rotgans, J. I., & Schmidt, H. G. (2017). The relation between individual interest and knowledge acquisition. *British Educational Research Journal*, 43(2), 350–371.
- Sánchez, E., García-Rodicio, H., & Acuña, S. R. (2009). Are instructional explanations more effective in the context of an impasse? *Instructional Science*, 37(6), 537–563.
- Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, 26(3–4), 299–323.
- Schiefele, U., & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and Individual Differences*, 8(2), 141–160.
- Schmidt, H. G., De Volder, M. L., De Grave, W. S., Moust, J. H. C., & Patel, V. L. (1989). Explanatory models in the processing of science text: The role of prior knowledge activation through small-group discussion. *Journal of Educational Psychology*, 81(4), 610–619.
- Schunk, D. H. (1995). Self-efficacy, motivation, and performance. *Journal of Applied Sport Psychology*, 7(2), 112–137.
- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, 16(4), 475–522.
- Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases. The effects of telling first on learning and transfer. *Journal of Educational Psychology*, 103(4), 759–775.
- Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction*, 22(2), 129–184.
- Sitzmann, T., & Yeo, G. (2013). A meta-analytic investigation of the within-person self-efficacy domain. Is self-efficacy a product of past performance or a driver of future performance? *Personnel Psychology*, 66(3), 531–568.
- Sweller, J., Kirschner, P., & Clark, R. E. (2007). Why minimally guided teaching techniques do not work: A reply to commentaries. *Educational Psychologist*, 42(2), 115–121.
- van Harsel, M., Hoogerheide, V., Verkoeijen, P., & van Gog, T. (2019). Effects of different sequences of examples and problems on motivation and learning. *Contemporary Educational Psychology*, 58, 260–275.
- VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21(3), 209–249.
- Vosniadou, S. (Ed.). (2013). *International handbook of research on conceptual change. Educational psychology handbook series* (2nd ed.). Routledge.
- Vosniadou, S., Lawson, M. J., Wyra, M., Van Deur, P., Jeffries, D., & Ngurah, D. I. (2020). Pre-service teachers' beliefs about learning and teaching and about the self-regulation of learning: A conceptual change perspective. *International Journal of Educational Research*, 99, 101495.
- Weaver, J. P., Chastain, R. J., DeCaro, D. A., & DeCaro, M. S. (2018). Reverse the routine. Problem solving before instruction improves conceptual knowledge in undergraduate physics. *Contemporary Educational Psychology*, 52, 36–47.
- Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In C. M. Wittrock (Ed.), *Handbook of research in teaching* (pp. 315–327). Macmillan Publishing Company.