



Example-based learning: should learners receive closed-book or open-book self-explanation prompts?

Sara Hiller^{1,3} · Stefan Rumann² · Kirsten Berthold³ · Julian Roelle¹ 

Received: 23 April 2019 / Accepted: 11 August 2020 / Published online: 1 September 2020
© The Author(s) 2020

Abstract

In learning from examples, students are often first provided with basic instructional explanations of new principles and concepts and second with examples thereof. In this sequence, it is important that learners self-explain by generating links between the basic instructional explanations' content and the examples. Therefore, it is well established that learners receive self-explanation prompts. However, there is hardly any research on whether these prompts should be provided in a closed-book format—in which learners cannot access the basic instructional explanations during self-explaining and thus have to retrieve the main content of the instructional explanations that is needed to explain the examples from memory (i.e., retrieval practice)—or in an open-book format in which learners can access the instructional explanations during self-explaining. In two experiments, we varied whether learners received closed- or open-book self-explanation prompts. We also varied whether learners were prompted to actively process the main content of the basic instructional explanations before they proceeded to the self-explanation prompts. When the learners were not prompted to actively process the basic instructional explanations, closed-book prompts yielded detrimental effects on immediate and delayed (1 week) posttest performance. When the learners were prompted to actively process the basic instructional explanations beforehand, closed-book self-explanation prompts were not less beneficial than open-book prompts regarding performance on a delayed posttest. We conclude that at least when the retention interval does not exceed 1 week, closed-book self-explanation prompts do not entail an added value and can even be harmful in comparison to open-book ones.

Keywords Example-based learning · Self-explanations · Generative learning activities · Retrieval practice · Prompts

✉ Julian Roelle
julian.roelle@ruhr-uni-bochum.de

¹ Faculty of Philosophy and Educational Research, Ruhr University Bochum, Universitätsstraße 150, 44801 Bochum, Germany

² Department of Chemistry Education, University of Duisburg-Essen, Schützenbahn 70, 45127 Essen, Germany

³ Department of Psychology, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany

Example-based learning is a common and powerful instructional means to introduce learners to new content (e.g., Hoogerheide and Roelle 2020; Van Gog et al. 2019). One highly effective sequence of learning from examples works as follows (see Renkl 2014; Wittwer and Renkl 2010). Firstly, learners are provided instructional explanations that communicate basic knowledge concerning new principles and concepts. Secondly, learners are given examples that illuminate these principles and concepts (e.g., Atkinson 2002; Foster et al. 2018; Schworm and Renkl 2007; Roelle et al. 2017). One important activity in learning from this sequence is generating *principle-based self-explanations*. During this learning activity, students explain the examples by using the content of the basic instructional explanations (e.g., Hausmann and VanLehn 2010; Renkl 2014).

As learners usually do not engage in this learning activity to a sufficient degree of their own accord, one important instructional principle for example-based learning is to encourage learners to self-explain the examples (Renkl 2014). Frequently, this principle is implemented by providing the examples in conjunction with *self-explanation prompts* (e.g., Atkinson et al. 2003; Conati and VanLehn 2000; Schalk et al. 2020). The benefits of these prompts are evident—a wealth of research clearly shows that self-explanation prompts foster the benefits of example-based learning (for an overview, see Renkl 2014). Nevertheless, self-explanation prompts are not always equally effective, which indicates that their benefits depend on certain factors.

One factor that could substantially moderate the benefits of self-explanation prompts but has scarcely been considered is the *format* of the self-explanation prompts. Basically, self-explanation prompts can be implemented in two different formats: a *closed-book format* in which learners cannot reinspect the basic instructional explanations while generating their self-explanations or an *open-book format*, in which learners have access to the basic instructional explanations while they self-explain the examples. In the present study, we investigated the role of this potential moderator. We predicted that when the retention interval is short, open-book self-explanation prompts would be more beneficial than closed-book prompts and that this effect would be mitigated when the retention interval is long. Furthermore, we predicted that closed-book self-explanation prompts might be even more effective than open-book self-explanation prompts when learners actively processed the main content items of the basic instructional explanations before they engaged in self-explaining and thus are well-prepared to respond to closed-book self-explanation prompts.

Below, we describe the theoretical rationale behind our predictions and then report two experiments. In both experiments, learners were introduced to new content by means of example-based learning. We manipulated (a) whether learners received closed- or open-book self-explanation prompts, and (b) whether learners were prompted to actively process the main content items of the basic instructional explanations before they received the self-explanation prompts. As the main dependent variables, we used the number of generated self-explanations as well as learners' performance on an immediate and a one-week delayed posttest.

Self-explanation prompts viewed through the lens of generative learning theory

Generative learning activities involve learners manipulating given information or generating new information in order to understand the information that has already been given to them (e.g., Fiorella and Mayer 2016). These activities are theorized to benefit learning because they

foster the degree to which coherent and elaborated mental representations of new content are formed and integrated with learners' prior knowledge (see Fiorella and Mayer 2016; Weinstein and Mayer 1986). The beneficial effects of generative activities are evident; a wealth of research has demonstrated that engaging in generative learning activities can significantly enhance learning outcomes (e.g., explanation tasks, see Hoogerheide et al. 2019; concept mapping tasks; see Lachner et al. 2017; drawing tasks, see Schmidgall et al. 2018; adjunct questions, see Roelle et al. 2014; learning protocols, see Nückles et al. 2020).

The process of self-explaining examples qualifies as a generative learning activity. Specifically, in generating principle-based self-explanations, which is considered a crucial learning activity in learning from examples (Renkl 2014), learners establish relations between the basic instructional explanations, which are given in the initial step of the example-based learning sequence, and the examples, which are provided in the second step of the sequence (e.g., Hausmann and VanLehn 2010; Hefter et al. 2015; Nokes et al. 2011). Viewed through the lens of generative learning theory, generating such principle-based self-explanations fosters learning via two mechanisms. First, it enhances the degree of elaboration of learners' mental representations of the basic instructional explanations' content. The underlying mechanism that leads to this enhancement of the mental representations is that in generating the self-explanations learners link the basic content (i.e., principles and concepts) to concrete examples (see Renkl 2014). At the same time, it enhances learners' mental representations of the examples. This is because self-explaining enhances the degree to which the examples are stored correspondent to the principles and concepts they reflect (i.e., their structural and relevant features; Reimann 1997).

Against this background, it is reasonable to conceive of self-explanation prompts as activators of generative learning activities. Viewed through the lens of generative learning theory, it furthermore is reasonable to predict that prompts designed to elicit self-explanations generally are more effective when they are implemented in an *open-book format* rather than in a *closed-book format*. A closed-book format raises the hurdle for learners to successfully establish the prompted relations between the basic instructional explanations and the examples because it presupposes that learners are able to successfully retrieve the content of the basic instructional explanations that is required to explain the concrete examples from memory. As retrieval is rarely perfect (e.g., Rowland 2014), a closed-book format likely reduces the amount of generated self-explanations as compared to an open-book format (see Roelle and Renkl 2020). From the perspective of generative learning theory, this hurdle is raised without good cause. The number of generated self-explanations is considered crucial, whereas the amount of retrieval that is involved in generating the self-explanations is not.

However, although it is the predominant perspective in the example-based learning literature, generative learning theory is not the only perspective that can be informative for understanding and optimizing the benefits of self-explanation prompts. At least in terms of understanding the learning activities that are elicited by closed-book self-explanation prompts, taking the perspective of retrieval-based learning theory can be fruitful as well.

Self-explanation prompts viewed through the lens of retrieval-based learning theory

The literature on retrieval-based learning shows that the act of retrieval from memory (i.e., *retrieval practice*) is an efficient way to foster learning (see Adesope et al. 2017; Roediger and Butler 2011; Rowland 2014). It indicates that even when time on task is controlled for

and when no feedback is provided, learners who practice the retrieval of specific idea units from memory show lower forgetting rates and thus demonstrate better long-term retention than learners who do not. The beneficial effects of retrieval practice are frequently explained by the fact it contributes to the consolidation of memory traces. Two main mechanisms that likely support this consolidation effect are outlined in the *episodic context account* and the *elaborative retrieval account*.

The episodic context account proposes that retrieving knowledge items leads to consolidation of the items' representation in memory because the context representation that is stored with the respective items is updated (Karpicke et al. 2014). In order to retrieve certain idea units or knowledge items, learners can reinstate the episodic context in which idea units were initially encoded. When learners are successful in retrieving the idea units, features from the context in which the idea units are currently retrieved are added to the context representation. These newly added context features can be used as further cues to access the idea units in memory on future retrieval occasions. Thus, through the addition of contextual features, the strength of the idea units' mental representation is enhanced. In contrast, the elaborative retrieval account (e.g., Carpenter 2009, 2011) proposes that in using certain retrieval cues to search for targeted idea units, several other idea units that are semantically related to the retrieval cues are activated and, as a consequence, more strongly linked to the representation of the targeted idea units. On future occasions these more strongly linked units can serve as additional retrieval routes for the respective idea units.

The beneficial effects of retrieval practice are well established. Numerous studies have indicated that engaging in (different types of) retrieval practice fosters learning (e.g., Bae et al. 2018; Butler 2010; Grimaldi and Karpicke 2014; Karpicke and Blunt 2011; Roediger and Karpicke 2006; Rummer et al. 2017). In example-based learning, self-explanation prompts can serve as retrieval practice tasks. More specifically, when self-explanation prompts are provided in a closed-book format, which does not allow learners to review the basic instructional explanations while responding to the prompts, the self-explanation prompts inevitably require learners to retrieve the main idea units of the basic instructional explanations required to explain the examples from memory.

When self-explanation prompts are viewed not only through the lens of generative learning theory but also through the lens of retrieval-based learning theory, at first glance it is reasonable to predict that closed-book self-explanation prompts should be more beneficial than open-book prompts. The rationale behind this prediction is that closed-book prompts engage learners in two types of beneficial learning activities (generative and retrieval practice activities), whereas open-book prompts engage learners only in one type of beneficial learning activity (generative activities). However, recent studies have indicated that this line of argumentation might be too simple.

The downsides of closed-book self-explanation prompts

Learners seldomly retrieve information perfectly (e.g., Dunlosky and Rawson 2015; Rowland 2014). Hence, it is reasonable to assume that learners who receive closed-book self-explanation prompts would not be able to retrieve all of the required knowledge components of the instructional explanations. This, in turn, should hinder the generation of self-explanations. As a result, one could predict that closed-book self-explanation prompts elicit fewer self-explanations than open-book self-explanation prompts. In line with this notion, Blunt and Karpicke (2014) reported that the learners in their study generated fewer

idea units when concept mapping and summarizing were prompted in a closed- rather than in an open-book format. Notably, this effect occurred even though the closed-book learners were provided with the opportunity to restudy the learning material at one point during concept mapping. Similarly, Roelle and Berthold (2017) found that implementing adjunct questions designed to prompt learners to elaborate on an expository text in a closed-book format led to a decrease in the number of generated elaborations in comparison to an open-book format (for similar results, see Agarwal et al. 2008; Roelle and Nückles 2019; Waldeyer et al. 2020). Hence, although learning tasks in which generative learning activities and retrieval practice are combined can benefit learning (e.g., Endres et al. 2017; Hinze et al. 2013; Heitmann et al. 2018), in comparison to *pure* generative learning tasks the combination of generative learning activities and retrieval practice entails certain downsides.

The downsides that come along with imperfect retrieval concern not only the *generative part* of learning tasks in which generative learning and retrieval practice is combined but also the *retrieval practice part*. At least when no feedback is provided, the benefits of practicing retrieval substantially depend on the degree to which learners are able to correctly retrieve the required knowledge items (e.g., Karpicke et al. 2014; Rowland 2014). Hence, the benefits of both the generative and the retrieval part of learning tasks should be impaired when learners are not able to successfully retrieve all of the required knowledge items from memory.

Against this background, it can be assumed that closed-book self-explanation prompts would not necessarily be more beneficial than open-book self-explanation prompts. However, given that retrieval practice decreases forgetting rates (see Rowland 2014), it can be predicted that the effects between closed-book and open-book self-explanation prompts change over time. When the retention interval is short (e.g., when a posttest follows immediately after the learning phase), open-book self-explanation prompts might be more beneficial than closed-book prompts because open-book prompts elicit more self-explanations and potential retrieval-driven benefits of closed-book prompts should be relatively low. This potential superiority of open-book prompts should decrease and might even disappear when the retention interval is long because the retrieval-driven benefits of closed-book self-explanation prompts might increasingly outweigh the lower number of generated self-explanations. These hypotheses are supported by previous studies that investigated effects between closed- and open-book generative learning tasks. Both Agarwal et al.'s (2008) and Roelle and Berthold's (2017) findings suggest that closed-book learning tasks catch up to open-book learning tasks over time.

However, at least when complex generative learning tasks (such as self-explanation prompts) were involved, these previous findings also show that a closed-book format did not lead to higher learning outcomes than an open-book format even after a delay (but see Rummer et al. 2019). One potential underlying reason for this lack of superiority of the closed-book format could be the large detrimental effect of a closed-book format regarding the number of elicited generative learning activities. This detrimental effect, however, should substantially depend on learners' processing in the initial study phase (i.e., before they engage in retrieval-based learning activities). With respect to example-based learning, this notion means that the degree to which learners actively process the main content items of the basic instructional explanations (provided in the initial step of the sequence) is crucial for the benefits of closed-book self-explanation prompts. Learners who actively process the main content of the basic instructional explanations should be able to successfully retrieve the required knowledge components when they receive the closed-book self-explanation prompts to a greater extent than learners who do not actively process the main content items beforehand (e.g., Fiorella and Mayer 2016; Kintsch et al. 1990). High retrieval

rates, in turn, should increase the benefits of both the retrieval practice part (because retrieval is effective only when it is successful) and the generative part of the self-explanation prompts (because learners are able to generate more self-explanations). It follows that if learners are encouraged to actively process the basic instructional explanations before they receive the closed-book self-explanation prompts, the benefits of a closed-book format should substantially increase; in this case, because they engage learners in (additional) retrieval practice without substantially hindering the generative activity of self-explaining, they might even be more effective than open-book self-explanation prompts, which solely engage learners in the generative activity of self-explaining, regarding performance on a delayed posttest.

Hypotheses

In view of the outlined findings and theoretical considerations, the present study was designed to investigate the role of the format (closed-book vs. open-book) of self-explanation prompts in example-based learning. In terms of learning processes, we hypothesized that when learners were *not* prompted to actively process the main content items of the basic instructional explanations before they receive the examples and self-explanation prompts, closed-book self-explanation prompts would elicit fewer self-explanations than open-book self-explanation prompts (hypothesis 1a). Under these circumstances, we furthermore predicted that open-book self-explanation prompts would yield higher learning outcomes on an immediate posttest (hypothesis 2a) but not on a delayed posttest (hypothesis 3a).

When the learners were prompted to actively process the main content items of the basic instructional explanations in the initial study phase, we predicted a different pattern of results. In this case, we did not expect a significant difference between closed- and open-book prompts regarding the number of generated self-explanations (hypothesis 1b) and immediate posttest performance (hypothesis 2b). Regarding delayed posttest performance, we expected the closed-book prompts to be more effective (hypothesis 3b).

Experiment 1

Method

Sample and design

The participants were $N=97$ eighth-grade students from German high schools (57 female). The students were between 12 and 15 years old ($M=13.43$, $SD=0.53$) and received €10 for their participation.¹ The first language of all participants was German. As the learning material was mainly text-based (see below), participants whose first language was not German were excluded from the experiment.

¹ The relatively large range regarding age was due to the fact that one student (15 years old) was kept back a year and thus repeated the eighth grade and one student (12 years old) had skipped a grade. Ninety-five students were either 13 or 14 years old.

The participants were introduced to three new topics via example-based learning. Specifically, for each topic the learners in the first step received a basic instructional explanation that was designed to impart basic knowledge concerning new principles and concepts. These basic instructional explanations were provided either with or without prompts that were designed to elicit active processing of the main content of the basic instructional explanations (hereafter: *active processing of instructional explanation prompts*). In the second step, all learners were provided with two examples of the principles and concepts that were communicated by the basic instructional explanations and two self-explanation prompts that asked the learners to explain the examples using the idea units of the basic instructional explanations. Depending on the experimental condition, the self-explanation prompts were provided in either a *closed-book format* in which the learners could not review the respective basic instructional explanation while generating their self-explanations or in an *open-book format* in which the basic instructional explanations were still available while the learners engaged in self-explaining.

Jointly, these manipulations resulted in a 2×2 factorial between-subjects design with the factors *active processing of instructional explanation prompts* (with vs. without) and *format of self-explanation prompts* (closed-book vs. open-book). The participants were randomly assigned to the conditions.

Materials

A science educationalist (the second author) supervised the design of all materials. All materials were based on, and thus highly similar to, the materials used in Roelle et al. (2017).

The computer-based learning environment covered three new topics that were part of the participants' regular curriculum: (a) the Bohr model, (b) the formation of ions (ionization), and (c) ionic reactions. The topics were introduced using the sequence that is established in the field of learning from worked examples (see Renkl 2014). Hence, each topic's introduction began with an instructional explanation providing basic declarative knowledge concerning the new concepts and principles. These basic instructional explanations were provided in text form and were between 163 and 192 words in length (see Fig. 1).

In the conditions with active processing of instructional explanation prompts, each of the basic instructional explanations was combined with a prompt that required the learners to actively process the basic instructional explanations' main content in their own words. More specifically, based on the recommendations by Berthold and Renkl (2010), the active processing of instructional explanations prompts were explicitly designed to focus learners on the most relevant content items, i.e. the content items that are crucial for self-explaining the subsequently provided examples. It is important to highlight that although the active processing of instructional explanations prompts asked the students to respond to the prompts in their own words (e.g., "Based on the text, answer the following question in your own words: How are atoms structured according to the Bohr model?"), they hardly required the learners to go beyond the content of the basic instructional explanations. Rather, the prompts mainly engaged learners in attending to the instructional explanations' main content items. Consequently, the active processing of instructional explanation prompts did not qualify as a type of self-explanation prompts, which usually require learners to go (substantially) beyond the provided information (e.g., by relating the content to one's prior knowledge or by explicating to what extent the respective information entails new insights, see e.g., Chi et al. 1994; Wylie and Chi 2014). The learners typed their prompts responses

Ionic Reactions

One way in which atoms of two elements can react with each other is the so-called ionic reaction. When an ionic reaction takes place, the atoms of two elements combine by one element giving up its electrons from its outermost shell to the outermost shell of another atom. Thus, ionic reactions involve the transfer of one or more electrons from the atom of one element to the atom of another element.

The following general principles hold true when ionic reactions occur:

- No electrons are lost when the reaction takes place. The same number of electrons are present both before and after the reaction. All of the electrons that are given up by an atom of one element are taken up by an atom of the other element.
- The transfer of electrons only occurs between the outermost shells of the atoms that are reacting to each other.
- Once an ionic reaction has taken place, all of the atoms involved have full outer shells.

For these reasons, ionic reactions do not always take place at a ratio of 1:1. Thus, it is possible that there is a higher number of atoms of one element than of the other element when ionic reactions take place.

Here you can see a couple of examples of ionic reactions. Sodium and fluoride always react with each other at a ratio of 1:1. One sodium atom reacts with one fluoride atom. By contrast, magnesium always reacts with fluoride at a ratio of 1:2. One magnesium atom reacts with two fluoride atoms.

Answer the following questions based on the basic principles ionic reactions you worked on above:

- Why does sodium react with fluoride at a ratio of 1:1?
- Why does magnesium react with fluoride at a ratio of 1:2?

a)

b)

continue

Ionic Reactions

One way in which atoms of two elements can react with each other is the so-called ionic reaction. When an ionic reaction takes place, the atoms of two elements combine by one element giving up its electrons from its outermost shell to the outermost shell of another atom. Thus, ionic reactions involve the transfer of one or more electrons from the atom of one element to the atom of another element.

The following general principles hold true when ionic reactions occur:

- No electrons are lost when the reaction takes place. The same number of electrons are present both before and after the reaction. All of the electrons that are given up by an atom of one element are taken up by an atom of the other element.
- The transfer of electrons only occurs between the outermost shells of the atoms that are reacting to each other.
- Once an ionic reaction has taken place, all of the atoms involved have full outer shells.

For these reasons, ionic reactions do not always take place at a ratio of 1:1. Thus, it is possible that there is a higher number of atoms of one element than of the other element when ionic reactions take place.

Here you can see a couple of examples of ionic reactions. Sodium and fluoride always react with each other at a ratio of 1:1. One sodium atom reacts with one fluoride atom. By contrast, magnesium always reacts with fluoride at a ratio of 1:2. One magnesium atom reacts with two fluoride atoms.

Answer the following questions based on the basic principles ionic reactions you worked on above:

- Why does sodium react with fluoride at a ratio of 1:1?
- Why does magnesium react with fluoride at a ratio of 1:2?

a)

b)

continue

Fig. 1 Screenshots of examples and self-explanation prompts in the closed-book (left screenshot) and open-book (right screenshot) conditions (translated from German). The screenshot of the open-book condition also includes a basic instructional explanation (upper part)

into text boxes that were positioned next to the basic instructional explanations. In the conditions without active processing of instructional explanation prompts, the students were given the instruction to “Please type any thoughts that come to mind into this text box as you attempt to understand the principles of the Bohr model/the formation of ions/ionic reactions.” Albeit this general prompt could be conceived as an active processing prompt as well, its stimulative nature likely was lower than that of the active processing of instructional explanation prompts.

After they had finished processing the basic instructional explanations, the learners proceeded to the two examples of the previously introduced basic content (i.e., principles and concepts). Each of these examples (between 44 and 73 words in length) included one graphic that was based on graphics that are typically used in high-school chemistry textbooks (see Fig. 1). Each example was provided together with a self-explanation prompt that asked the students to explain the respective example by using the content of the basic instructional explanations. For example, the learners had to “Answer the following question based on the principles of the formation of ions you worked on above: Why do sodium atoms form positive ions?”.

In the conditions in which the self-explanation prompts were provided in a closed-book format, the basic instructional explanations were hidden once the students proceeded to the examples and self-explanation prompts (see Fig. 1). Thus, the learners in these conditions had to retrieve the idea units that were needed to make sense of the examples from memory. By contrast, in the conditions in which the self-explanation prompts were provided in an open-book format, the basic instructional explanations were still presented on the screen when the learners proceeded to the examples. Thus, these learners were not dependent on retrieval from memory in generating self-explanations. In order to prevent test expectation effects (e.g., Agarwal and Roediger 2011), the learners were not informed about the format of self-explanation prompts in advance. The learners typed their prompts responses into text boxes that were positioned next to the examples. The copy and paste commands were disabled so that the learners had to actively type their text box entries.

Instruments and measures

Pretest: assessment of prior knowledge A pretest assessed the participants' prior knowledge using 10 open-ended questions. Four questions assessed very basic prior chemistry knowledge (e.g., "What is an electron?"). The other six items directly related to the three content topics of the learning environment. For instance, the learners were asked "How are atoms structured according to the Bohr model?" or "Why do sodium atoms form positive ions?".

Using a scoring protocol, two independent raters (blind to the conditions) scored the responses of 20 learners. For each of the questions, the number of correct arguments was determined (1 point per argument; incomplete but correct arguments were awarded 0.5 points). Interrater reliability, as determined by the intraclass coefficient (absolute agreement), was very good for each of the 10 questions (all ICC > .85). Thus, only one rater scored the rest of the written answers. For the later analyses, we summed up the points over all 10 questions (Cronbach's $\alpha = .79$).

Prompts responses: assessment of learning processes In order to gain insight into the learning processes the learners executed throughout the example-based learning sequence, we separately analyzed the entries in the two types of text boxes (i.e., the text boxes that were provided together with the basic instructional explanations and the text boxes that were provided together with the examples). Regarding instructional explanation processing, the overwhelming majority of learners' text box entries were comprised of restatements of the main content items that were included in the basic instructional explanations (e.g., facts such as "Each atom has a nucleus.", "All atoms have protons." and "Protons are inside the nucleus."). These content items were subsumed under the category *covered main content items*. There were also some segments in which the learners related the content of the basic instructional explanations to their prior knowledge (e.g., the atom "[...] wants to be moved to the state of a noble gas."). However, segments that fell into this category of *integration/elaboration* were very rare (< 1% of all segments).

In terms of example processing, most of the text box entries were comprised of relations between the basic instructional explanations and the examples that were generated by the learners during self-explaining. These segments were subsumed under the category of *self-explanations*. In some instances, the learners referred to knowledge that was not communicated by the basic instructional explanations in their self-explanations (e.g., "Sodium has three electron shells because it is in the third period."). As these references to prior knowledge were very rare (less than one reference per participant), these self-explanations were not differentiated from the self-explanations that solely referred to content of the basic instructional explanations.

Two independent raters analyzed the entries of 20 students. To distinguish between the different content items and interrelations, they used a list of separately coverable content items and interrelations for each basic instructional explanation and for each example. As the examples that were to be explained partly included multiple features that could be related to the basic instructional explanations (e.g., in explaining why sodium atoms react with fluoride atoms at a ratio of 1:1, the learners needed to refer to knowledge components regarding the occupation of electron shells and regarding principles of ionic reactions), it was possible that the learners generated more than one self-explanation per example. Specifically, each generated interrelation between an example and the previous basic instructional explanations that did not overlap with previously

generated interrelations was counted as a self-explanation. In view of the findings that both erroneous processing of basic instructional explanations (e.g., Roelle et al. 2015) and erroneous self-explanations (e.g., Berthold and Renkl 2009) can detrimentally affect learning outcomes, segments that included errors were not scored.² Interrater reliability was very good for both the basic instructional explanation and the example processing categories (all ICC > .85).

Posttest: assessment of learning outcomes A posttest assessed the students' conceptual knowledge using 18 open-ended questions (six items per topic). Six of the 18 questions were identical to questions used in the pretest, while the other 12 items were new. Four of the new questions were similar to the self-explanation prompts (e.g., "Why does beryllium form a positive ion?"), whereas the other eight came in new formats such as "This example about the formation of boron ions contains a mistake. Give reasons why this example is incorrect and explain how a boron ion is correctly formed."

Based on a scoring protocol that was analogous to the one used in the pretest, two independent raters (blind to the conditions) scored the answers of 20 students. Interrater reliability was very good for each of the 18 questions (all ICC > .85). For the later analyses, the points were summed up over all 18 questions (Cronbach's $\alpha_{\text{immediate}} = .84$; Cronbach's $\alpha_{\text{delayed}} = .87$).

Procedure

The experiment took place in the students' regular classrooms. Throughout the entire experiment, the students worked individually at a computer. First, they took the pretest. Second, they were provided the basic instructional explanations (with or without active processing of instructional explanation prompts) and examples with self-explanation prompts (closed-book or open-book format) regarding the three topics. All learners took a posttest immediately after the learning phase as well as one week later. After they had completed the immediate posttest, the learners were instructed not to deal with the learning content and not to talk to each other about the learning content until the end of the second session (delayed posttest). At the end of the second session, all learners indicated that they had complied with this instruction. The time spent on the posttest was limited to 36 min for both the immediate and the delayed posttest.

Results

Table 1 shows the mean scores and standard deviations for the four groups on all measures of the study. An α -level of .05 was used for all tests.

² The number of errors when processing both the instructional explanations and the examples were very low, $M = 0.65$ ($SD = 1.18$) and $M = 0.53$ ($SD = 0.85$), respectively. There were no statistically significant differences between the four conditions concerning these measures, $.57 > p > .14$.

Preliminary analyses

To analyze whether the random assignment had resulted in comparable groups, we first compared the students' pretest scores and chemistry grades (all students were in classes of high-track high schools (i.e., *Gymnasium* in Germany) and thus the approaches to award grades should be comparable among all participants). We did not find any statistically significant effects of condition, $F(3, 93)=0.22$, $p=.876$, $\eta_p^2=.00$, and $F(3, 93)=1.15$, $p=.331$, $\eta_p^2=.03$. Hence, the groups did not significantly differ concerning these important learning prerequisites. To reduce error variance, both variables were included as covariates in our subsequent analyses. The assumption of homogeneous within group regression slopes was not violated in any of the analyses.³

Learning processes

We were interested in whether closed-book self-explanation prompts would elicit fewer self-explanations than open-book self-explanation prompts when the learners were not prompted to actively process the main content of the basic instructional explanations before they proceeded to the examples and self-explanation prompts (hypothesis 1a). Furthermore, we sought to find out whether the potential detrimental effect of closed-book self-explanation prompts would be mitigated when learners were prompted to actively process the main content of the basic instructional explanations beforehand (hypothesis 1b).

In terms of covariates, we found a statistically significant effect of the pretest score but not of the chemistry grade, $F(1, 91)=32.11$, $p<.001$, $\eta_p^2=.26$, and $F(1, 91)=0.24$, $p=.620$, $\eta_p^2=.00$. Concerning main effects, we did not find a statistically significant effect of either the format of self-explanation prompts, $F(1, 91)=1.50$, $p=.219$, $\eta_p^2=.01$, or the active processing of instructional explanation prompts, $F(1, 91)=0.15$, $p=.694$, $\eta_p^2=.00$. However, there was a statistically significant interaction effect, $F(1, 91)=4.36$, $p=.040$, $\eta_p^2=.04$. The pattern of the interaction effect is shown in Fig. 2a. Probing the interaction revealed that the closed-book format reduced the number of generated self-explanations for the learners who did not receive active processing of instructional explanation prompts, $F(1, 44)=4.44$, $p=.041$, $\eta_p^2=.09$, but not for the learners with active processing of instructional explanation prompts, $F(1, 45)=0.40$, $p=.526$, $\eta_p^2=.00$.

We also analyzed whether the active processing of instructional explanation prompts actually fostered the degree to which the students actively processed the main content items of the basic instructional explanations before they received the examples and self-explanation prompts. Regarding the number of covered main content items, the ANCOVA

³ Please note that we did not conduct multilevel analyses because the participants were individually randomly assigned to the conditions and all dependent variables were measured on the individual level. Hence, even if the classrooms from which the participants stemmed would have differed concerning their performance, the pattern of results should not be biased due to the potential influence of the L2-variable classroom. Furthermore, before we addressed our main hypotheses, which were related to the dependent variables generated self-explanations and performance on the immediate and delayed posttest by conducting three separate ANCOVAs, we conducted a MANCOVA that included all three main dependent variables. This MANCOVA revealed statistically significant effects of both covariates, $F(3, 75)=6.78$, $p<.001$, $\eta_p^2=.21$ for chemistry grade, and $F(3, 75)=11.45$, $p<.001$, $\eta_p^2=.31$ for pretest score, but no statistically significant main effects, $F(3, 75)=1.79$, $p=.155$, $\eta_p^2=.06$ for the active processing of instructional explanation prompts, and $F(3, 75)=0.52$, $p=.667$, $\eta_p^2=.02$ for the format of self-explanation prompts. By contrast, there was a statistically significant interaction effect, $F(3, 75)=3.05$, $p=.034$, $\eta_p^2=.10$.

Table 1 Means and (standard deviations) of all measures in Experiment 1

	Active processing of instructional explanation prompts/open-book self-explanation prompts	No active processing of instructional explanation prompts/open-book self-explanation prompts	Active processing of instructional explanation prompts/closed-book self-explanation prompts	No active processing of instructional explanation prompts/closed-book self-explanation prompts
Pretest	12.51 (9.57)	12.50 (8.78)	10.96 (7.31)	11.25 (8.20)
Chemistry grade ^a	2.05 (0.78)	2.46 (0.83)	2.33 (0.87)	2.38 (0.64)
Learning processes				
Self-explanations	9.72 (4.15)	11.75 (5.27)	9.92 (4.73)	8.54 (5.18)
Incorrect self-explanations	0.68 (0.83)	0.54 (0.93)	0.29 (0.54)	0.66 (1.04)
Covered main content items	19.09 (6.53)	11.45 (10.45)	18.51 (5.82)	11.95 (9.44)
Integration/elaboration	0.04 (0.21)	0.16 (0.56)	0.18 (0.39)	0.08 (0.40)
Posttest	34.06 (13.28)	35.64 (12.78)	33.44 (13.59)	28.35 (11.41)

^aIn German high schools 1 is the best and 6 is the worst grade than can be attained

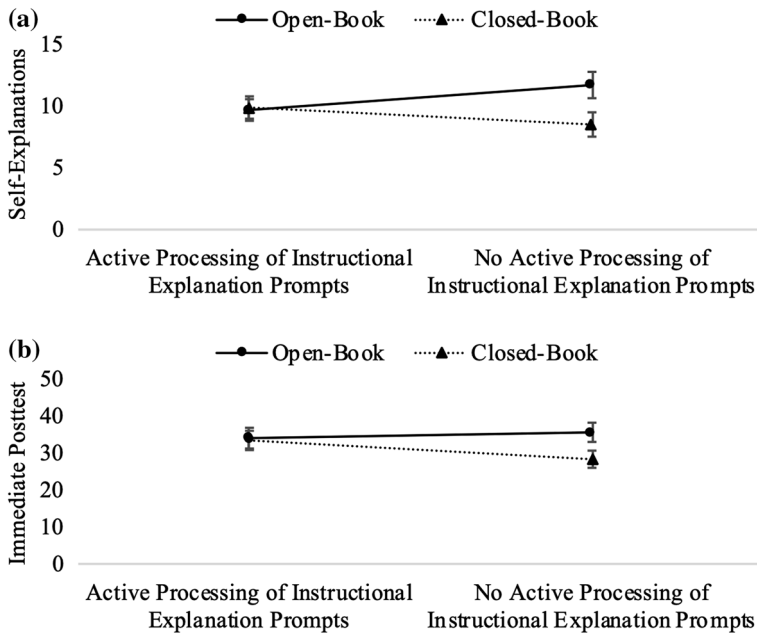


Fig. 2 Interactions between active processing of instructional explanation prompts (with vs. without) and format of self-explanation prompts (closed-book vs. open-book) regarding the number of self-explanations and immediate posttest scores in Experiment 1. Error bars represent standard errors of the means

revealed a statistically significant main effect of active processing of instructional explanation prompts, $F(1, 93) = 15.98$, $p < .001$, $\eta_p^2 = .14$. The learners who received the active processing of instructional explanation prompts covered the main content items to a higher extent in their text box entries than their counterparts. The covariate chemistry grade was a statistically significant predictor in this model, $F(1, 93) = 4.40$, $p = .039$, $\eta_p^2 = .04$, whereas the pretest was not a statistically significant predictor, $F(1, 93) = 2.62$, $p = .109$, $\eta_p^2 = .02$.

With respect to the degree to which the students related content items of the basic instructional explanations to their prior knowledge, there was no statistically significant effect of the active processing of instructional explanation prompts, $F(1, 93) = 0.00$, $p = .927$, $\eta_p^2 = .00$. Furthermore, none of the covariates entailed a statistically significant effect (both F s < 1).

Learning outcomes

We predicted that when learners were *not* prompted to actively process the main content items of the basic instructional explanations before they received the examples and self-explanation prompts, open-book self-explanation prompts would yield higher learning outcomes than closed-book self-explanation prompts on an immediate posttest (hypothesis 2a). Furthermore, we hypothesized that this effect would be mitigated when the posttest was delayed (hypothesis 3a). When the learners were prompted to actively process the main content of the basic instructional explanations in the initial study phase, we did not expect to find a significant effect of the format of prompts regarding immediate posttest

performance (hypothesis 2b). In terms of delayed posttest performance, we assumed that closed-book prompts would be more effective than open-book prompts (hypothesis 3b).

To address these hypotheses, we conducted a mixed repeated measures ANCOVA with time of posttest (immediate vs. delayed) as the within-subjects factor and active processing of instructional explanation prompts and format of self-explanation prompts as between-subjects factors. The ANCOVA revealed statistically significant effects for both covariates: $F(1, 77) = 22.20$, $p < .001$, $\eta_p^2 = .22$ for the pretest, and $F(1, 77) = 8.90$, $p = .004$, $\eta_p^2 = .10$ for the chemistry grade, respectively. Furthermore, on the between-subjects level we found neither a statistically significant main effect of active processing of instructional explanation prompts, $F(1, 77) = 0.32$, $p = .571$, $\eta_p^2 = .00$, nor a statistically significant main effect of format of self-explanation prompts, $F(1, 77) = 1.00$, $p = .319$, $\eta_p^2 = .01$. Also, there was no statistically significant interaction effect, $F(1, 77) = 2.34$, $p = .130$, $\eta_p^2 = .03$. None of these effects was qualified by an interaction with the measurement time point. The interaction between the measurement time point and active processing of instructional explanation prompts, $F(1, 77) = 3.24$, $p = .076$, $\eta_p^2 = .04$, the interaction between measurement time point and the format of self-explanation prompts, $F(1, 77) = 0.00$, $p = .949$, $\eta_p^2 = .00$, and the three-way interaction, $F(1, 77) = 1.88$, $p = .174$, $\eta_p^2 = .02$, did not reach statistical significance. There was, however, a statistically significant effect of the measurement time point, $F(1, 77) = 4.60$, $p = .035$, $\eta_p^2 = .05$. Not surprisingly, the immediate posttest scores were higher than the delayed posttest scores.

In view of the fact that 14 learners missed the delayed posttest (due to a sports event at one of the schools), which in the above-mentioned mixed ANCOVA reduces the statistical power not only regarding the delayed but also regarding the immediate posttest, we also conducted separate analyses for the immediate posttest and the delayed posttest. Regarding immediate posttest performance, we found statistically significant effects of both the covariate pretest, $F(1, 91) = 28.77$, $p < .001$, $\eta_p^2 = .24$, and the covariate chemistry grade, $F(1, 91) = 15.82$, $p < .001$, $\eta_p^2 = .14$. In terms of main effects, as in the mixed ANCOVA there was no statistically significant effect of the active processing of instructional explanation prompts, $F(1, 91) = 0.08$, $p = .772$, $\eta_p^2 = .00$, and also no statistically significant effect of the format of the self-explanation prompts, $F(1, 91) = 1.33$, $p = .251$, $\eta_p^2 = .01$. However, there was a small but statistically significant interaction effect, $F(1, 91) = 4.40$, $p = .039$, $\eta_p^2 = .04$. The pattern of the interaction is depicted in Fig. 2b. Probing the interaction revealed that the closed-book self-explanation prompts decreased posttest performance for the learners without active processing of instructional explanation prompts, $F(1, 44) = 5.69$, $p = .021$, $\eta_p^2 = .11$, but not for the learners with active processing of instructional explanation prompts, $F(1, 49) = 0.35$, $p = .552$, $\eta_p^2 = .00$.

For explorative purposes, in the next step we analyzed whether in the groups without active processing of instructional explanation prompts, the higher immediate posttest performance of the learners with the open-book self-explanation prompts was due to the fact that these learners had generated more self-explanations in the learning phase than their closed-book counterparts. In order to address this question, we performed a mediation analysis using Hayes' (2013) SPSS macro PROCESS. Specifically, we calculated 95% bootstrap percentile confidence intervals of the potential mediation effect from 10,000 bootstrap samples. The results of the mediation analysis are shown in Fig. 3. We found a statistically significant indirect effect via the generated self-explanations, $a \times b = 3.46$ [0.09, 8.31]. Hence, because of the mediating function of the generated self-explanations, the learners in the open-book group had an advantage of 3.46 units over the learners in the closed-book group on the immediate posttest.

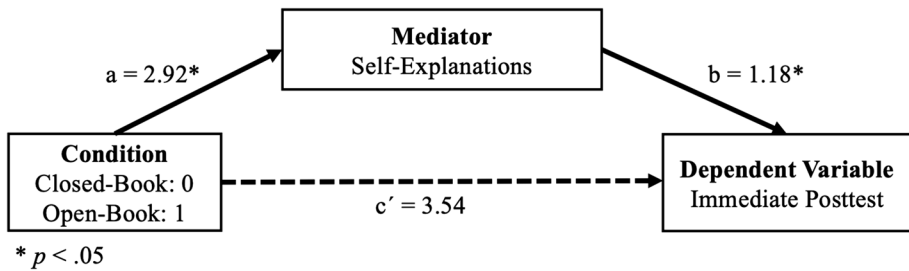


Fig. 3 Results of the mediation analysis in Experiment 1

In terms of delayed posttest performance, we found a slightly different pattern of results. Similar to the immediate posttest, both the pretest score and chemistry grade served as significant predictors, $F(1, 77) = 19.10$, $p < .001$, $\eta_p^2 = .19$, and $F(1, 77) = 4.92$, $p = .029$, $\eta_p^2 = .06$. Furthermore, similar to the immediate posttest, there were no statistically significant main effects, $F(1, 77) = 0.89$, $p = .347$, $\eta_p^2 = .01$ for the format of the self-explanation prompts, and $F(1, 77) = 0.77$, $p = .383$, $\eta_p^2 = .01$ for the active processing of instructional explanation prompts, respectively. However, other than at the immediate posttest, there was no statistically significant interaction effect, $F(1, 77) = 1.05$, $p = .307$, $\eta_p^2 = .01$.

Experiment 2

The findings of Experiment 1 indicate that closed-book self-explanation prompts can have detrimental effects as compared to open-book self-explanation prompts. When the learners were not prompted to actively process the main content items of the basic instructional explanations beforehand, the open-book self-explanation prompts were more beneficial in terms of both learning processes (generated self-explanations) and learning outcomes as measured by an immediate posttest (hypotheses 1a and 2a). However, both effects between the closed- and open-book self-explanation prompts decreased and did not reach statistical significance when the learners were prompted to actively process the main content items of the basic instructional explanations before proceeding to the examples and self-explanation prompts (hypotheses 1b and 2b). In terms of delayed learning outcomes, as expected we did not find a significant effect of the prompt format when the learners were not prompted to actively process the main content of the basic instructional explanations (hypothesis 3a). However, contrary to our expectation, the closed-book self-explanations prompts were not superior to the open-book prompts when the learners received active processing of instructional explanation prompts beforehand (hypothesis 3b). The effect sizes and lack of significant interaction effects of the mixed ANCOVA furthermore indicate that the differences concerning the pattern of results between the immediate and delayed posttest were relatively small. Thus, these results thus should be interpreted cautiously.

The results regarding the groups that were not prompted to actively process the main content items of the basic instructional explanations are in line with previous research that compared closed- and open-book generative learning tasks (Agarwal et al. 2008; Roelle and Berthold 2017; Waldeyer et al. 2020). More specifically, the mediation analysis indicates that because it detrimentally affects the number of executed generative learning activities (here: self-explanations), a closed-book format detrimentally affects immediate

posttest performance. In terms of delayed posttest performance, the detrimental effect of the closed-book format is substantially decreased; however, even in this case the closed-book format is not superior to an open-book format.

Our results with respect to the groups that were prompted to actively process the main content items of the basic instructional explanations go beyond previous findings. These results reveal that the detrimental effect of a closed-book format regarding the execution of generative learning activities can be reduced through encouraging learners to engage in active processing in the initial study phase (i.e., before learners engage in retrieval). However, even in this case we did not find a significant superiority of the closed-book prompts. At first glance, these findings question whether closed-book self-explanation prompts can be more beneficial than open-book self-explanation prompts in example-based learning. However, the pattern of results regarding delayed posttest performance should be interpreted with caution because it was confounded by the presence of the immediate posttest. Specifically, the immediate posttest can be conceived of as a further retrieval practice activity, which itself affected learning outcomes. As the groups differed regarding their performance on the immediate posttest, it likely biased the pattern of results of the delayed posttest.

Against this background, in Experiment 2 we pursued two main goals. First, in light of recent emphasis on the importance of replicating novel findings (e.g., Maner 2014; Simons 2014), we aimed at replicating the results of Experiment 1 regarding learning processes (generation of self-explanations) with a second sample of eighth-grade high school students. Second, we aimed at testing the prediction that closed-book self-explanation prompts would be superior to open-book self-explanation prompts when the posttest was delayed and when learners were prompted to actively process the main content items of the basic instructional explanations beforehand (hypothesis 3b) without confounding through an immediate posttest.

Method

Sample and design

We recruited $N=124$ eighth-grade students from different German high schools (88 female) as participants for Experiment 2 (as in Experiment 1, students whose first language was not German were excluded from the experiment). They were between 12 and 14 years old ($M=13.24$, $SD=0.44$) and received €10 for their participation.⁴ As in Experiment 1, the learners were randomly assigned to one condition of a 2×2 factorial between-subjects design with the factors *active processing of instructional explanation prompts* (with vs. without) and *format of self-explanation prompts* (closed-book vs. open-book).

Materials

We used the same learning materials as in Experiment 1.

⁴ The relatively large range regarding age was due to the fact that one student (12 years old) had skipped a grade. The majority of the students ($N=123$) were either 13 or 14 years old.

Instruments and measures

Pretest: assessment of prior knowledge The pretest was identical to Experiment 1. Two raters scored the answers of 20 students (all ICC > .85). In view of the high interrater reliability, only one rater scored the rest of the written answers. The points were summed up over all 10 questions for the later analyses (Cronbach's $\alpha = .71$).

Prompts responses: assessment of learning processes The students' text box entries were examined using the same procedure as in Experiment 1 with the only exception that the category *elaboration/integration* was not used for instructional explanation processing because there were no segments that fell into this category.⁵ Interrater reliability was very good for both the instructional explanation and the example processing (all ICC > .85).

Posttest: assessment of learning outcomes The posttest was identical to the one used in Experiment 1. Interrater reliability was very good for each of the 18 questions (all ICC > .85). For the later analyses, the points were summed up over all 18 questions (Cronbach's $\alpha = .80$).

Procedure

Except for the fact that there was no immediate posttest, the procedure was identical to Experiment 1. At the end of the first session, the learners were instructed not to deal with the learning content and not to talk to each other about the learning content until the end of the second session. At the end of the second session, all learners indicated that they had complied with this instruction. Due to illness, 10 participants missed the delayed posttest.

Results

Table 2 shows the mean scores and standard deviations for the four groups on all measures of the study. An α -level of .05 was used for all tests.

Preliminary analyses

Prior to addressing our hypotheses, we first tested whether there were differences between the groups with regard to their pretest scores and chemistry grades. In terms of both variables, we did not find a statistically significant effect of condition, $F(3, 120) = 1.22$, $p = .302$, $\eta_p^2 = .03$, and $F(3, 120) = 1.24$, $p = .287$, $\eta_p^2 = .03$. Thus, there were no significant differences between the groups with respect to these important learning prerequisites. As in Experiment 1, both pretest score and chemistry grade were included as covariates in

⁵ As in Experiment 1, the number of errors during both processing the instructional explanations and processing the examples was very low, $M = 0.75$ ($SD = 1.14$) and $M = 0.68$ ($SD = 0.98$), respectively. There were no statistically significant differences between the four conditions concerning these measures, $.96 > p > .30$.

Table 2 Means and (standard deviations) of all measures in Experiment 2

	Active processing of instructional explanation prompts/open-book self-explanation prompts	No active processing of instructional explanation prompts/open-book self-explanation prompts	Active processing of instructional explanation prompts/closed-book self-explanation prompts	No active processing of instructional explanation prompts/closed-book self-explanation prompts
Pretest	2.67 (3.31)	4.53 (6.10)	3.04 (4.44)	4.25 (3.97)
Chemistry grade	2.19 (0.98)	1.93 (0.78)	1.97 (0.60)	2.26 (0.81)
Learning processes				
Self-explanations	5.77 (4.61)	8.16 (4.51)	6.12 (5.77)	4.77 (4.79)
Incorrect self-explanations	0.61 (0.84)	0.81 (0.92)	0.58 (0.99)	0.74 (1.14)
Covered main content items	18.83 (4.34)	10.18 (9.73)	20.74 (4.91)	13.08 (12.25)
Posttest	17.12 (10.56)	16.56 (8.02)	21.10 (12.72)	10.09 (7.43)

the subsequent analyses. For all analyses, the assumption of homogeneous within group regression slopes was not violated.⁶

Learning processes

As in Experiment 1, we were interested in whether closed-book self-explanation prompts would elicit fewer self-explanations than open-book self-explanation prompts when the learners were not prompted to actively process the main content items of the basic instructional explanations before they proceeded to the examples and self-explanation prompts (hypothesis 1a). Moreover, we sought to find out whether the potential detrimental effect of closed-book self-explanation prompts would be mitigated when learners received active processing of instructional explanation prompts beforehand (hypothesis 1b).

In terms of covariates, the ANCOVA revealed neither a statistically significant effect of pretest, $F(1, 118)=0.45$, $p=.500$, $\eta_p^2=.00$, nor of chemistry grade, $F(1, 118)=0.50$, $p=.481$, $\eta_p^2=.00$. Concerning main effects, there was no statistically significant main effect of the format of self-explanation prompts, $F(1, 118)=2.76$, $p=.099$, $\eta_p^2=.02$, and also no statistically significant main effect of active processing of instructional explanation prompts, $F(1, 118)=0.47$, $p=.493$, $\eta_p^2=.00$. However, there was a marginally significant interaction effect, $F(1, 118)=3.83$, $p=.053$, $\eta_p^2=.03$. Although the interaction effect was merely marginally significant, for explorative purposes we analyzed whether the interaction pattern (see Fig. 4a) corresponded with the pattern that was found in Experiment 1. For the learners without active processing of instructional explanation prompts, the closed-book self-explanations significantly decreased the number of self-explanations, $F(1, 58)=7.65$, $p=.008$, $\eta_p^2=.11$; this effect was not found for the learners with active processing of instructional explanation prompts, $F(1, 58)=0.01$, $p=.903$, $\eta_p^2=.00$.

We also analyzed whether and to what extent the active processing of instructional explanation prompts increased the degree to which the learners actively processed the main content items of the basic instructional explanations before they received the examples and self-explanation prompts. An ANCOVA showed that the active processing of instructional explanation prompts substantially increased the extent to which the learners covered the main content items of the basic instructional explanations in their text box entries, $F(1, 120)=26.27$, $p<.001$, $\eta_p^2=.18$. The covariates pretest score and chemistry grade were not statistically significant predictors in this model (both F s < 1).

Learning outcomes

Regarding learning outcomes, we assumed that when learners were *not* prompted to actively process the main content items of the basic instructional explanations before they receive the examples and self-explanation prompts, there would be no significant difference

⁶ Similar to Experiment 1, before we addressed our main hypotheses, which were related to the dependent variables generated self-explanations and performance on the delayed posttest by conducting separate ANCOVAs, we conducted a MANCOVA that included both main dependent variables. This MANCOVA revealed no statistically significant effects of the covariates, $F(2, 107)=0.64$, $p=.526$, $\eta_p^2=.01$ for chemistry grade, and $F(2, 107)=0.33$, $p=.720$, $\eta_p^2=.00$ for pretest score. Further, there was a statistically significant main effect of active processing of instructional explanations prompts, $F(2, 107)=12.71$, $p<.001$, $\eta_p^2=.19$, but not of the format of self-explanation prompts, $F(2, 107)=1.62$, $p=.202$, $\eta_p^2=.02$. The interaction effect between the two factors was statistically significant, $F(2, 107)=3.52$, $p=.033$, $\eta_p^2=.06$.

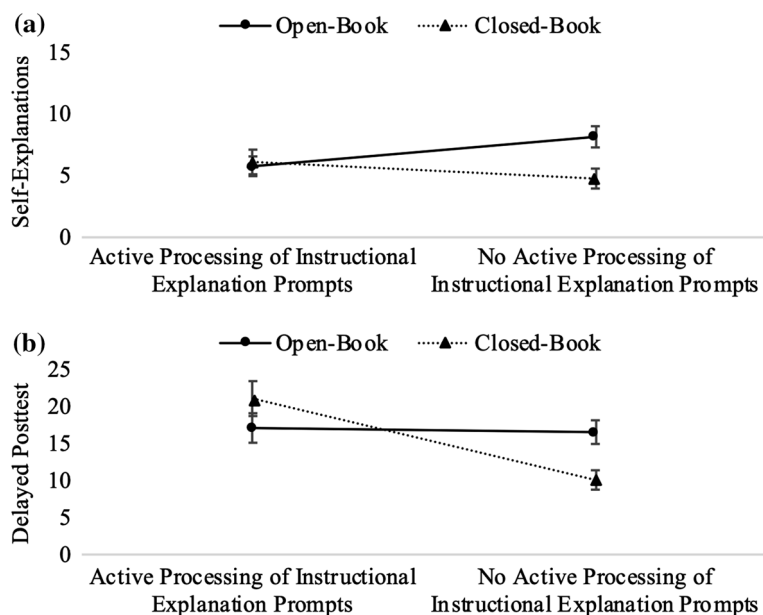


Fig. 4 Interactions between active processing of instructional explanation prompts (with vs. without) and format of self-explanation prompts (closed-book vs. open-book) regarding the number of self-explanations and delayed posttest scores in Experiment 2. Error bars represent standard errors of the means

between open- and closed-book self-explanation prompts in terms of delayed posttest performance (hypothesis 3a). However, when the learners were prompted to actively process the main content items of the basic instructional explanations in the initial study phase, we assumed that closed-book prompts would be superior (hypothesis 3b).

With respect to covariates, the ANCOVA indicated neither a statistically significant effect of the chemistry grade, $F(1, 108)=0.05$, $p=.809$, $\eta_p^2=.00$, nor of the pretest score, $F(1, 108)=0.65$, $p=.420$, $\eta_p^2=.00$. Regarding main effects, we found a statistically significant effect of active processing of instructional explanation prompts, $F(1, 108)=8.99$, $p=.003$, $\eta_p^2=.07$. The learners who received active processing of instructional explanation prompts achieved higher scores than the learners who did not receive active processing of instructional explanation prompts. There was no statistically significant main effect of the format of self-explanation prompts, $F(1, 108)=0.33$, $p=.563$, $\eta_p^2=.00$. However, there was a statistically significant interaction effect, $F(1, 108)=7.10$, $p=.009$, $\eta_p^2=.06$. The pattern of the interaction effect is shown in Fig. 4b. Probing the interaction showed that the interaction was due to the fact that the closed-book self-explanation prompts significantly decreased posttest performance for the learners without active processing of instructional explanation prompts, $F(1, 53)=11.85$, $p=.001$, $\eta_p^2=.18$, but not for the learners with active processing of instructional explanation prompts, $F(1, 53)=1.45$, $p=.232$, $\eta_p^2=.02$.

General discussion

In the present study, we investigated the role of the format (closed-book vs. open-book) of self-explanation prompts in example-based learning. We predicted that closed-book self-explanation prompts would hinder self-explanations when learners were not prompted to actively process the main content items of the basic instructional explanations before they received the examples and self-explanation prompts, which should detrimentally affect immediate but not necessarily delayed posttest performance. When the learners were prompted to actively process the main content items of the basic instructional explanations in the initial study phase, we did not predict a significant difference between closed- and open-book prompts regarding self-explanations and immediate posttest performance. Rather, in this case we predicted a superiority of the closed-book self-explanation prompts concerning delayed posttest performance.

Closed- vs. open-book self-explanation prompts: effects on learning processes

When there were no prompts that elicited active processing of the main content items of the basic instructional explanations that were provided in the first phase of the example-based learning sequence, closed-book self-explanation prompts reduced the number of generated principle-based self-explanations in both experiments (hypothesis 1a). One explanation for this result is that the learners who received the closed-book self-explanation prompts were not able to successfully retrieve all of the required knowledge components of the basic instructional explanations when they engaged in self-explaining. Hence, in comparison to the learners in the open-book groups who could review the basic instructional explanations and thus were not dependent on successful and complete retrieval while self-explaining, the closed-book learners were at a disadvantage when it came to establishing interrelations between the basic instructional explanations and examples. This pattern of results is consistent with previous findings regarding the effects of the format (open-book vs. closed-book) of generative learning tasks (see Agarwal et al. 2008; Blunt and Karpicke 2014; Roelle and Berthold 2017; Waldeyer et al. 2020).

In both experiments, we also found that the detrimental effect of the closed-book self-explanation prompts was substantially mitigated when the learners were prompted to actively process the main content items of the basic instructional explanations that were provided in the first step of the example-based learning sequence (hypothesis 1b). An explanation for this finding, which goes beyond previous research, is that the active processing that was elicited by the prompts enhanced the quality of the mental representations that the learners formed when processing the basic instructional explanations. Based on generative learning theory, it is reasonable to assume that these enhanced mental representations fostered learners' retention and understanding of the main content of the basic instructional explanations (e.g., Fiorella and Mayer 2016; Kintsch et al. 1990; Roelle and Nückles 2019). Hence, by fostering active processing, the prompts decreased the forgetting rates of the knowledge components of the basic instructional explanations which, in turn, helped the closed-book learners to generate nearly the same number of principle-based self-explanations as their open-book counterparts. Notably, these consistent results were found even though the samples in Experiment 1 and 2 substantially differed in terms of prior knowledge. Although implicitly already reflected in the fact that the assumption of homogeneous within group regression slopes was not

violated in any of the analyses for the covariate pretest score, this result suggests that prior knowledge does not seem to be a substantial moderator of the effects of the format of self-explanation prompts in example-based learning.

It is important to highlight, however, that the (marginally) significant interactions regarding the generation of self-explanations that were found in both experiments likely were due in part to an unexpected side-effect of the active processing of instructional explanation prompts for the learners with open-book self-explanation prompts. More specifically, an inspection of the interaction patterns in both experiments suggests that the active processing of instructional explanation prompts detrimentally affected the generation of self-explanations on part of the open-book learners (see Figs. 2a, 4a). One explanation for this pattern of results could be that responding to the active processing of instructional explanation prompts was exhausting, which might have led the learners who received open-book self-explanation prompts to respond to the self-explanation prompts in a relatively economic manner. Consequently, in comparison to their counterparts without active processing of instructional explanation prompts, they did not exploit the full potential of the open-book self-explanation prompts.

Against this background, the lack of effect between the closed- and open-book self-explanations prompts for the groups that received active processing of instructional explanation prompts beforehand should be interpreted with caution. Future studies should test whether active processing of instructional explanation prompts would also be sufficient to prevent detrimental effects of closed-book self-explanation prompts when learners are allowed to take breaks in order to recover from the potential exhaustion that is due to responding to active processing of instructional explanation prompts. In these future studies, it would furthermore be useful to use larger sample sizes and conduct a priori power analyses, which were lacking in the present study. In terms of the low to medium interaction effects regarding generated self-explanations, the power of our experiments to detect these effects was relatively low, which reduces the interpretability of the respective findings.

Closed- vs. open-book self-explanation prompts: effects on learning outcomes

In terms of learning outcomes, our findings did not fully support our hypotheses. More specifically, in line with hypothesis 2a we found that when the learners were not prompted to actively process the main content items of the basic instructional explanations that were provided in the first step, the closed-book self-explanation prompts yielded lower performance on the immediate posttest than the open-book self-explanation prompts. However, contrary to hypothesis 3a, this inferiority of the closed-book prompts was not consistently mitigated when the posttest was delayed; at least when there was no confounding through an immediate posttest, the closed-book learners were outperformed by the open-book learners regarding delayed posttest performance (Experiment 2).

One explanation for this pattern of results is that the engagement in retrieval practice on part of the closed-book learners did not fall on fertile ground. As these learners generated fewer self-explanations than their counterparts, their mental representations of the learning content likely were of relatively low quality after the learning phase. Consolidating these deficient mental representations might have entailed relatively little benefit regarding learning outcomes, which resulted in an inferiority of the closed-book learners even at the delayed posttest. An alternative explanation could be that the retention interval that was used in the present study was too short. For instance, Rummer et al.'s (2017; see also

Rummer et al. 2019) results suggest that when retrieval practice tasks are compared to relatively strong control conditions (such as an open-book equivalent), their benefits might be found only after longer delays (e.g., after two weeks). Hence, in future studies it might be fruitful to test whether closed-book self-explanation prompts—even when they are not preceded by active processing of instructional explanation prompts—might entail beneficial effects in comparison to open-book prompts after longer retention intervals.

When the learners were prompted to actively process the main content of the basic instructional explanations beforehand, we found a different pattern of results. Specifically, when there was no confounding through an immediate posttest, we found that the closed-book group was not outperformed by the open-book group regarding delayed posttest performance in Experiment 2 (hypothesis 3b). One explanation for this pattern of results is that the active processing of instructional explanation prompts prevented the closed-book learners from generating fewer self-explanations than their open-book counterparts. Consequently, the quality of the mental representations after the learning phase should have been equal for the closed- and the open-book learners. Yet even under these circumstances, the additional retrieval on part of the closed-book learners scarcely paid off. As mentioned above, one explanation for why the benefit of the closed-book format was relatively small and did not reach statistical significance even when the closed-book learners did not generate fewer self-explanations than their open-book counterparts could be that the retention interval was too short. A further explanation could be that the delay between processing the basic instructional explanations and responding to the self-explanation prompts was too short as well. In the retrieval-based learning literature, there is evidence which suggests that the benefits of engaging in retrieval practice increase with an increasing delay between the initial study phase and the retrieval phase because the temporal context hardly changes when the delay is short. Consequently, little context updating takes place and thus the process of retrieval hardly contributes to building distinctive context cues that can be used to retrieve the information in the future (see Karpicke et al. 2014). Hence, in future studies it could be useful to analyze the benefits of closed-book self-explanation prompts in settings with longer delays between processing the basic instructional explanations and self-explaining the examples.

In addition to exploring the outlined potential means to enhance the benefits of closed-book self-explanation prompts, it could also be fruitful to investigate potential optimizations of open-book self-explanations prompts as well. In the present study, although the copy and paste commands were disabled, the open-book learners could nevertheless exactly type or closely paraphrase content items of the basic instructional explanations. In view of the finding that paraphrasing is less beneficial than engaging in deep-oriented generative learning activities (e.g., Hausmann and VanLehn 2010), a consequence of the learners' engagement in exactly typing and paraphrasing could be that they did not exploit the full potential of the open-book self-explanation prompts. Informing learners about the low effectiveness of copying and paraphrasing during self-explaining could be a viable means to reduce their engagement in these activities and could thus further enhance the benefits of open-book self-explanation prompts. Future studies should address this potential optimization of open-book self-explanation prompts.

A further limitation that should be addressed in future research relates to the prompts that were designed to elicit active processing of the basic instructional explanations' main content. As stated in the Method section of Experiment 1, these prompts were merely designed to engage learners in attending to the basic instructional explanations' main content items but scarcely required learners' to deeply engage with the provided content.

Consequently, the learners in the groups with and without active processing of instructional explanations prompts differed only concerning the quantity of covered content items but not concerning the types of learning processes, for the learners without the active processing of instructional explanation prompts engaged in active processing of the instructional explanations content as well to some degree (i.e., ca. 10–13 covered idea units, see Tables 1 and 2). In view of this pattern of results and the finding that prompts that require learners to generate inferences or elaborations and thus go beyond the provided information can be substantially more effective than active processing prompts that merely require attending processes (e.g., Chi 2009; Roelle et al. 2015), our active processing of instructional explanations prompts likely were suboptimal. Although the prompts were sufficient to substantially decrease the detrimental effect of closed-book self-explanation prompts concerning the generation of self-explanations, it would be interesting to test whether the example-based learning sequence that was used in the present study could be further optimized by increasing the depth of the processing of the basic instructional explanations.

Conclusions

The format of self-explanation prompts clearly matters. When learners are not encouraged to actively process the main content items of the basic instructional explanations before they receive the examples and self-explanation prompts, open-book self-explanation prompts elicit more self-explanations than closed-book self-explanation prompts, which can beneficially affect both immediate posttest performance (Experiment 1) and delayed posttest performance (Experiment 2). By contrast, when learners are encouraged to actively process the main content of the basic instructional explanations beforehand, open-book and closed-book self-explanation prompts do not significantly differ concerning the number of elicited self-explanations. However, even in this case open-book self-explanation prompts are not less beneficial than closed-book self-explanation prompts concerning learning outcomes, at least when the retention interval does not exceed one week. We conclude that instructors should provide learners with open-book prompts for self-explanation because in case of sufficient processing of the basic instructional explanations they are not less beneficial and in case of insufficient processing of the basic instructional explanations they are more beneficial than closed-book self-explanation prompts.

Acknowledgements The research in this article was funded by the *Deutsche Forschungsgemeinschaft* (German Research Foundation; DFG, GZ: RO 5012/1-1). We would like to thank the following research assistants: Alica Buchau, Marius Schäfer, Paula Scheibe, and Lisa Sieweke for their assistance in conducting the experiments, coding the process data, and analyzing the tests, Alexander Meerhoff for helping with computer programming, and Stewart Campbell for proofreading. Furthermore, we thank all the students who participated in our studies as well as all teachers for their support.

Funding The research in this article was funded by the *Deutsche Forschungsgemeinschaft* (German Research Foundation; DFG, GZ: RO 5012/1-1). Open Access funding provided by Projekt DEAL.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest.

Ethical approval The experiments were approved by the ethics committee of Bielefeld University (No. 2015 – 105).

Informed consent Written informed parental consent was given 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

- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, 87, 659–701. <https://doi.org/10.3102/0034654316689306>.
- Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22, 861–876. <https://doi.org/10.1002/acp.1391>.
- Agarwal, P. K., & Roediger, H. L. (2011). Expectancy of an open-book test decreases performance on a delayed closed-book test. *Memory*, 19, 836–852. <https://doi.org/10.1080/09658211.2011.613840>.
- Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, 94, 416–427. <https://doi.org/10.1037/0022-0663.94.2.416>.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Combining fading with prompting fosters learning. *Journal of Educational Psychology*, 95, 774–783. <https://doi.org/10.1037/0022-0663.95.4.774>.
- Bae, C. L., Theriault, D. J., & Redifer, J. L. (2018). Investigating the testing effect: Retrieval as a characteristic of effective study strategies. *Learning and Instruction*, 60, 206–214. <https://doi.org/10.1016/j.learninstruc.2017.12.008>.
- Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology*, 101, 70–87. <https://doi.org/10.1037/a0013247>.
- Berthold, K., & Renkl, A. (2010). How to foster active processing of explanations in instructional communication. *Educational Psychology Review*, 22, 25–40. <https://doi.org/10.1007/s10648-010-9124-9>.
- Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology*, 106, 849–858. <https://doi.org/10.1037/a0035934>.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1118–1133. <https://doi.org/10.1037/a0019902>.
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1563–1569. <https://doi.org/10.1037/a0017021>.
- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1547–1552. <https://doi.org/10.1037/a0024140>.
- Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1, 73–105. <https://doi.org/10.1111/j.1756-8765.2008.01005.x>.
- Chi, M., deLeeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439–477. [https://doi.org/10.1016/0364-0213\(94\)90016-7](https://doi.org/10.1016/0364-0213(94)90016-7).
- Conati, C., & VanLehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, 11, 398–415.
- Dunlosky, J., & Rawson, K. A. (2015). Do students use testing and feedback while learning? A focus on key concept definitions and learning to criterion. *Learning and Instruction*, 39, 32–44. <https://doi.org/10.1016/j.learninstruc.2015.05.003>.

- Endres, T., Carpenter, S., Martin, A., & Renkl, A. (2017). Enhancing learning by retrieval: Enriching free recall with elaborative prompting. *Learning and Instruction*, 49, 13–20. <https://doi.org/10.1016/j.learninstruc.2016.11.010>.
- Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review*, 28, 717–741. <https://doi.org/10.1007/s10648-015-9348-9>.
- Foster, N. L., Rawson, K. A., & Dunlosky, J. (2018). Self-regulated learning of principle-based concepts: Do students prefer worked examples, faded examples, or problem solving? *Learning and Instruction*, 55, 124–138. <https://doi.org/10.1016/j.learninstruc.2017.10.002>.
- Grimaldi, P. J., & Karpicke, J. D. (2014). Guided retrieval practice of educational materials using automated scoring. *Journal of Educational Psychology*, 106, 58–68. <https://doi.org/10.1037/a0033208>.
- Hausmann, R. G. M., & VanLehn, K. (2010). The effect of self-explaining on robust learning. *International Journal of Artificial Intelligence in Education*, 20, 303–332. <https://doi.org/10.3233/JAI-2010-010>.
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.
- Hefter, M. H., Renkl, A., Riess, W., Schmid, S., Fries, S., & Berthold, K. (2015). Effects of a training intervention to foster precursors of evaluativist epistemological understanding and intellectual values. *Learning and Instruction*, 38, 11–22. <https://doi.org/10.1016/j.learninstruc.2015.05.002>.
- Heitmann, S., Grund, A., Berthold, K., Fries, S., & Roelle, J. (2018). Testing is more desirable when it is adaptive and still desirable when compared to note-taking. *Frontiers in Psychology*, 9, 2596. <https://doi.org/10.3389/fpsyg.2018.02596>.
- Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory and Language*, 69, 151–164. <https://doi.org/10.1016/j.jml.2013.03.002>.
- Hoogerheide, V., Renkl, A., Fiorella, L., Paas, F., & Van Gog, T. (2019). Enhancing example-based learning: Teaching on video increases arousal and improves problem-solving performance. *Journal of Educational Psychology*, 111, 45–56. <https://doi.org/10.1037/edu0000272>.
- Hoogerheide, V., & Roelle, J. (2020). Example-based learning: New theoretical perspectives and use-inspired advances to a contemporary instructional approach. *Applied Cognitive Psychology*, 34, 787–792. <https://doi.org/10.1002/acp.3706>.
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331, 772–775. <https://doi.org/10.1126/science.1199327>.
- Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. Ross (Ed.), *The psychology of learning and motivation* (pp. 234–287). New York: Elsevier.
- Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language*, 29, 133–159. [https://doi.org/10.1016/0749-596X\(90\)90069-C](https://doi.org/10.1016/0749-596X(90)90069-C).
- Lachner, A., Burkhart, C., & Nückles, M. (2017). Mind the gap! Automated concept map feedback supports students in writing cohesive explanations. *Journal of Experimental Psychology: Applied*, 23, 29–46. <https://doi.org/10.1037/xap0000111>.
- Maner, J. K. (2014). Let's put our money where our mouth is. If authors are to change their ways, reviewers (and editors) must change with them. *Perspectives in Psychological Science*, 9, 343–351. <https://doi.org/10.1177/1745691614528215>.
- Nokes, T. J., Hausmann, R. G. M., VanLehn, K., & Gershman, S. (2011). Testing the instructional fit hypothesis: the case of self-explanation prompts. *Instructional Science*, 39, 645–666. <https://doi.org/10.1007/s11251-010-9151-4>.
- 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*. <https://doi.org/10.1007/s10648-020-09541-1>.
- Reimann, P. (1997). *Lernprozesse beim Wissenserwerb aus Beispielen [Learning processes of knowledge acquisition from examples]*. Bern, CH: Huber.
- Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science*, 38, 1–37. <https://doi.org/10.1111/cogs.12086>.
- Roediger, H. L., & Butler, C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15, 20–27. <https://doi.org/10.1016/j.tics.2010.09.003>.
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>.
- Roelle, J., & Berthold, K. (2017). Effects of incorporating retrieval into learning tasks: The complexity of the tasks matters. *Learning and Instruction*, 49, 142–156. <https://doi.org/10.1016/j.learninstruc.2017.01.008>.

- Roelle, J., Berthold, K., & Renkl, A. (2014). Two instructional aids to optimise processing and learning from instructional explanations. *Instructional Science*, 42, 207–228. <https://doi.org/10.1007/s11251-013-9277-2>.
- Roelle, J., Hiller, S., Berthold, K., & Rumann, S. (2017). Example-based learning: The benefits of prompting organization before providing examples. *Learning and Instruction*, 49, 1–12. <https://doi.org/10.1016/j.learninstruc.2016.11.012>.
- Roelle, J., Lehmkuhl, N., Beyer, M.-U., & Berthold, K. (2015a). The role of specificity, targeted learning activities, and prior knowledge for the effects of relevance instructions. *Journal of Educational Psychology*, 107, 705–723. <https://doi.org/10.1037/edu0000010>.
- Roelle, J., Müller, C., Roelle, D., & Berthold, K. (2015b). Learning from instructional explanations: Effects of prompts based on the active-constructive-interactive framework. *PLoS ONE*, 10(4), e0124115. <https://doi.org/10.1371/journal.pone.0124115>.
- Roelle, J., & Nückles, M. (2019). Generative learning versus retrieval practice in learning from text: The cohesion and elaboration of the text matters. *Journal of Educational Psychology*, 111, 1341–1361. <https://doi.org/10.1037/edu0000345>.
- Roelle, J., & Renkl, A. (2020). Does an option to review instructional explanations enhance example-based learning? It depends on learners' academic self-concept. *Journal of Educational Psychology*, 112, 131–147. <https://doi.org/10.1037/edu0000365>.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, 140, 1432–1463. <https://doi.org/10.1037/a0037559>.
- Rummer, R., Schweppe, J., Gerst, K., & Wagner, S. (2017). Is testing a more effective learning strategy than note-taking? *Journal of Experimental Psychology: Applied*, 23, 293–300. <https://doi.org/10.1037/xap0000134>.
- Rummer, R., Schweppe, J., & Schwede, A. (2019). Open-book versus closed-book tests in university classes: A field experiment. *Frontiers in Psychology*, 10, 463. <https://doi.org/10.3389/fpsyg.2019.00463>.
- Schalk, L., Roelle, J., Saalbach, H., Berthold, K., Stern, E., & Renkl, A. (2020). Providing worked examples for learning multiple principles. *Applied Cognitive Psychology*, 34, 813–824. <https://doi.org/10.1002/acp.3653>.
- Schmidgall, S. P., Eitel, A., & Scheiter, K. (2018). Why do learners who draw perform well? Investigating the role of visualization, generation and externalization in learner-generated drawing. *Learning and Instruction*, 60, 138–153. <https://doi.org/10.1016/j.learninstruc.2018.01.006>.
- Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, 99, 285–296. <https://doi.org/10.1037/0022-0663.99.2.285>.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9, 76–80. <https://doi.org/10.1177/1745691613514755>.
- Van Gog, T., Rummel, N., & Renkl, A. (2019). Learning how to solve problems by studying examples. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 183–208). New York, NY: Cambridge University Press. <https://doi.org/10.1017/9781108235631.009>.
- Waldeyer, J., Heitmann, S., Moning, J., & Roelle, J. (2020). Can generative learning tasks be optimized by incorporation of retrieval practice? *Journal of Applied Research in Memory and Cognition*. <https://doi.org/10.1016/j.jarmac.2020.05.001>.
- 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). New York: Macmillan Publishing Company.
- Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review*, 22, 393–409. <https://doi.org/10.1007/s10648-010-9136-5>.
- Wylie, R., & Chi, M. T. H. (2014). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 413–432). New York, NY: Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.021>.