REVIEW ARTICLE

Happy Together? On the Relationship Between Research on Retrieval Practice and Generative Learning Using the Case of Follow-Up Learning Tasks

Julian Roelle[1] · Tino Endres[2] · Roman Abel[1] · Niklas Obergassel[1] · Matthias Nückles[3] · Alexander Renkl[2]

Abstract

Generative learning activities are assumed to support the construction of coherent mental representations of to-be-learned content, whereas retrieval practice is assumed to support the consolidation of mental representations in memory. Considering such functions that complement each other in learning, research on how generative learning and retrieval practice intersect appears to be very fruitful. Nevertheless, the relationship between these two fields of research is "expandable"—research on generative learning and retrieval practice has been pursued so far largely side by side without taking much note of each other. Against this background, the present article aims to give this relationship a boost. For this purpose, we use the case of follow-up learning tasks provided after learners have processed new material in an initial study phase to illustrate how these two research strands have already inspired each other and how they might do so even more in the future. In doing so, we address open- and closed-book formats of follow-up learning tasks, sequences of follow-up learning tasks that mainly engage learners in generative activities and tasks that mainly engage learners in retrieval practice, and discuss commonalities and differences between indirect effects of retrieval practice and generative learning activities. We further highlight what we do and do not know about how these two activity types interact. Our article closes with a discussion on how the relationship between generative learning and retrieval practice research could bear (more and riper) fruit in the future.

Keyword Generative learning · Retrieval practice · Test-enhanced learning · Comprehension · Retention

		Degree of generative activities required	
		Low	*High*
Degree of retrieval required	*Low*	Open-book tasks not requiring that learners make sense of the material (i.e., tasks designed to elicit neither generative activities nor retrieval practice) *Examples:* Restudy tasks Rereading tasks	Open-book tasks requiring learners to make sense of the material (i.e., tasks designed to elicit mainly generative activities) *Examples:* Generative drawing or prompted self-explaining that can be performed with accessible learning material
	High	Closed-book tasks not requiring that learners make sense of the material (i.e., tasks designed to elicit mainly retrieval practice) *Examples:* Free recall tasks or quiz questions that tap factual knowledge	Closed-book tasks requiring learners to make sense of the material (i.e., tasks designed to elicit both generative activities and retrieval practice) *Examples:* Generative drawing or prompted self-explaining that are performed with the learning material not accessible

Fig. 1 Possible combinations of the degree of required generative activities and degree of required retrieval as key features of follow-up learning tasks

There are two features of follow-up learning tasks (i.e., learning tasks provided after learners have encountered new content in an initial study phase) that can substantially affect learning outcomes. These are the degree to which the tasks require that learners *make sense of the provided information* and the degree to which the tasks require that learners *retrieve the studied information from memory*. Making sense of provided information is closely related to generative learning activities such as organization and elaboration. These activities are assumed to increase the degree to which coherent mental representations are constructed that are well-integrated with learners' prior knowledge (e.g., Fiorella, 2023; Fiorella & Mayer, 2016). Retrieving knowledge from memory is essentially a core component of *retrieval practice* or *test-enhanced learning*. Retrieval practice is assumed to consolidate the respective retrieved knowledge in memory and hence enable learners to maintain their knowledge or performance originating from that knowledge over long periods of time (e.g., Karpicke, 2017; Roediger & Butler, 2011).

In designing follow-up learning tasks, these two features can be manipulated independent of each other (see Fig. 1). The degree to which engaging in generative activities is required can be manipulated through the task instructions (e.g., low degree: reread the learning material, high degree: generate your own examples illustrating the learning content). The degree to which retrieving knowledge

from memory is required can be manipulated by providing or not providing learners with access to the learning material while they work on the follow-up learning task (i.e., low degree: open-book format, high degree: closed-book format, see Agarwal et al., 2008).

There is a wealth of evidence that adding one of the two task features can substantially enhance learning outcomes in comparison to follow-up learning tasks that neither engage learners in sense-making through organization and elaboration activities nor require learners to retrieve initially studied information from memory (i.e., *restudy tasks*) (for recent overviews concerning the generative component, see Brod, 2021; Fiorella & Mayer, 2016; for recent overviews concerning the retrieval component, see Agarwal et al., 2021; Carpenter et al., 2022; Yang et al., 2021). By contrast, far less is known about the effects of combining these two features—be it the combination within one learning task or sequential combination across different learning tasks. This research gap might be partially due to the fact that contemporary research paradigms and questions associated with the two task features originate from different communities (educational psychology vs. cognitive psychology). Such origins from diverging research fields have led to substantial differences in the type of learning material (complex conceptual materials vs. relatively simple factual materials), the type of learning outcomes (comprehension and transfer tests vs. recall tests), and the timing of their measurement (immediate vs. delayed) predominantly used in the respective studies.

In this context, it is not surprising that the research on how generative learning and retrieval-based learning intersect is "expandable." Nevertheless, getting insight into the effects of combinations of the two features would be very interesting from both a theoretical perspective (the functions of the two features could complement each other) and practical perspective (authentic learning tasks often combine these features). Although some working groups have recently investigated and discussed if, when, and why it makes sense to combine both task features (e.g., Endres et al., 2017; Hinze et al., 2013; see also McDaniel, 2023; Roelle et al., 2022b) and although the differences outlined in terms of learning material and outcome measures keep getting smaller, research on generative learning and on retrieval practice research has so far been pursued largely side by side, taking little note of each other.

The present article aims to strengthen the relationship between generative learning and retrieval practice research. For this purpose, by relying on the case of follow-up learning tasks, we illustrate how these two strands of research have already inspired each other, albeit to a limited extent, and show how they might inspire each other even more in the future. In doing so, we first introduce briefly the generative learning research and retrieval practice research ecospheres. Next, we describe research on follow-up learning tasks designed to elicit generative activities that vary the degree to which engaging in retrieval practice is required. Likewise, we address research on follow-up learning tasks designed to elicit retrieval practice and that also vary the degree to which generative activities are required. We also discuss the value of directly comparing the effects of the two task features and illustrate why exploring the sequences of tasks eliciting mainly generative activities and tasks eliciting mainly retrieval practice would have added value. Furthermore, we discuss how indirect effects of tasks designed to mainly elicit retrieval practice (i.e., the effects

of retrieval practice on learning activities revealed *after* the retrieval) and indirect effects of tasks designed to elicit mainly generative activities can be sensibly combined. In conclusion, we highlight how the relationship between generative learning and retrieval practice research might bear (more and riper) fruit in the future.

It is important to highlight that in addressing the aforementioned issues, we take a mainly *educational perspective*. That is, we are more interested in *instructional effects* of the (interplay of the) two task features (e.g., regarding learning outcomes) than in commonalities and differences in the *cognitive mechanisms* driving the effects of the (interplay of the) two features. It is also important to emphasize that both task features are of course almost always present to some extent. In (almost) every follow-up task designed to engage learners in generative activities, the retrieval of previously studied information from memory is required to some extent, and in (almost) every learning task designed to engage learners in retrieval from memory, organization, elaboration, or inference generation are also triggered somewhat. Learning tasks can nevertheless differ significantly in the extent to which the respective activities are stimulated. The present article focuses on the effects of and open questions about such learning tasks.

The Generative Learning and Retrieval Practice Research Ecospheres

In generative learning research, effects of different degrees to which learning tasks require learners to engage in sense making of the learning material is key. The specific sense-making activities triggered can differ among diverse types of generative learning tasks (for a recent framework conceptualizing these differences, see Fiorella, 2023). On a general level, however, the benefits of different generative learning activities can be explained by the same general mechanism, whether it be generative drawing (for a recent overview, see Fiorella & Zhang, 2018), creating one's own illustrative examples (e.g., Froese & Roelle, 2022, 2023; Rawson & Dunlosky, 2016), or the generation of explanations (e.g., Lachner et al., 2022). Learners engage in organization (e.g., identifying main ideas and their interrelations), elaboration (e.g., connecting new information to prior knowledge and prior experiences by finding own examples or analogies and reformulating in own words), and inference generation (e.g., determining the consequences of certain states of affairs). According to various generative learning theories, such activities foster the coherence of learners' mental representations of the to-be-learned content and the integration of these mental representations with prior knowledge (e.g., Chi & Wylie, 2014; Fiorella & Mayer, 2016).[1]

---

[1] Please note that the present concept of generative learning goes beyond the classical *generation effect*, that is, the finding that the active generation of parts of to-be-learned information, typically word associates, fosters learning better than rereading (e.g., Bertsch et al., 2007; Slamecka & Graf, 1978; for studies investigating this effect in learning from text, see e.g., Abel & Hänze, 2019; Schindler & Richter, 2023). The generative activities of organization, elaboration, and inference generation do not necessarily mean that learners generate parts of the to-be-learned information, but that they restructure it.

In principle, tasks engaging learners in generative activities can be provided before learners receive initial instruction on a topic or during or after an initial learning phase (see Fiorella, 2023). In the present article, we focus on generative learning tasks provided *after* learners were initially instructed (e.g., by reading an expository text or attending a classroom lesson) and hence after learners have formed an initial mental representation of the learning content, so that we can better compare them to the typical retrieval practice tasks usually also undertaken after initial instruction. Note, however, that such provision after an initial study phase does *not* necessarily mean that the learning material is no longer present when learners execute generative activities. Rather, generative learning activities elicited after an initial study phase can be elicited in open-book (i.e., learning material is available) and closed-book format (i.e., learning material is not available).

A prototypical investigation of effects of the degree to which follow-up learning tasks require generative activities is as follows: in an initial study phase, learners read an expository text or study multimedia learning material introducing new content to be understood on a conceptual level (e.g., the Doppler Effect, see Fiorella & Mayer, 2013; potential causes of endocarditis, see Lachner et al., 2021; the biological process of influenza, see Schmeck et al., 2014). Then, learners are assigned either a task requiring generative activities (e.g., to draw pictures reflecting the main elements and relations described in a text or to generate own examples illustrating main concepts) or a task demanding hardly any generative activities such as restudy. Ideally, the degree to which the learner is required to retrieve previously studied information from memory (i.e., closed-book or open-book format) does not vary between the two tasks. Immediately after learners have performed the tasks, they take a posttest in which retention and, most importantly, conceptual knowledge and transfer are tested. The typical finding would be that learners whose tasks incorporated more generative activities outperform their counterparts with respect to conceptual knowledge and transfer in particular (e.g., Fiorella & Mayer, 2016; for boundary conditions of the benefits of generative learning, see e.g., Brod, 2021).

However, a prototypical study in the ecosphere of retrieval practice research has a different design. There is also an initial study phase, but learners would encode relatively simple factual content in this phase (e.g., vocabulary lists or brief prose passages, about 250–300 words, on topics like sea otters or the sun, see Roediger & Karpicke, 2006). Such content need not be or cannot be understood on a conceptual level. Once the learners have processed the content, they are given either a task requiring retrieval of the studied content from memory (e.g., the task to write down everything they can remember from the initial study phase or quiz questions tapping specific facts) or a task not (or hardly) requiring retrieval of the content such as restudy. These tasks may be repeated several times with interspersed phases while learners can reinspect the learning material. Ideally, the degree to which engaging in generative activities is required does not vary between the two tasks. There may be a posttest immediately afterwards, but in any case, there is at least one, usually several, time-delayed posttests (e.g., days or weeks later). These posttests mainly assess retention, though tasks requiring relatively near transfer (e.g., Butler, 2010) could be included as well. The typical finding is that learners assigned tasks demanding more retrieval practice outperform their counterparts in terms of long-term retention, but

not necessarily on posttests given immediately after the learning phase (for meta-analyses, see Adesope et al., 2017; Rowland, 2014; Yang et al., 2021). These benefits are largest when the format of the posttest questions matches the answer formats of the follow-up learning task (e.g., short answer or multiple-choice format in both phases, see Agarwal et al., 2021).

First Meeting Points of Generative Learning and Retrieval Practice Research

Although previous research on generative learning or retrieval practice has focused mainly on research within its own ecosphere, several studies have triggered substantial interest in both worlds. For instance, from the perspective of research on generative learning, a critical incident during the *retrieval practice research wave* as reflected in the exponential growth in publications since Roediger and Karpicke's (2006) seminal study (see Karpicke, 2017; Yang et al., 2021) is obvious in Karpicke and Blunt's (2011) provocative study. Their study compared the effects of a free recall task (i.e., a task designed to elicit mainly retrieval practice, see Fig. 1) and an open-book concept mapping task (i.e., a task designed to elicit mainly generative learning, see Fig. 1). Their main finding was that free recall substantially outperformed open-book concept mapping in terms of learning outcomes, which motivated researchers in the field of generative learning to investigate the concept of retrieval-based learning more energetically. One lesson learned from this closer inspection of the retrieval-based learning concept by generative learning researchers was that in many tasks established as *generative learning tasks*, such as journal writing (see Nückles et al., 2020), self-explaining worked examples (see Roelle & Renkl, 2020), learning by teaching (see Lachner et al., 2022), or adjunct questions (i.e., questions typically provided while or after learners read textbook material, see Hamaker, 1986), retrieval practice might be an overlooked active ingredient contributing to the beneficial effects on learning outcomes. In many implementations of these tasks, learners were required to engage in generative activities (e.g., generating an example while writing a learning journal entry or generating an explanation while learning by teaching) without direct access to the respective learning material (i.e., closed-book format). Hence, the learners needed to retrieve information from memory before they could engage in generative activities, making the tasks at least in part to retrieval practice tasks as well.

As a result, how exactly the effects of these alleged *generative learning tasks* are interpreted had to be questioned in many cases. It was no longer uncontroversial to attribute the benefits of these tasks on learning outcomes exclusively to the contribution of generative activities to forming coherent, elaborate mental representations, as suggested by the frequent theoretical explanations of the benefits of tasks engaging learners in generative activities (Fiorella & Mayer, 2016; see also Chi & Wylie, 2014; Wittrock, 2010). Rather, the benefits could be at least partially due to the consolidation of existing mental representations (i.e., the mental representations formed during the initial study phase) through retrieval practice as well. These potential consolidation benefits could be explained by spreading activation

in *elaborative retrieval* (i.e., semantically related knowledge items become associated with the retrieved knowledge)[2] or by *episodic context updating* (more different episodic context features become associated with the retrieved knowledge, making it easier to access on future occasions). A discussion of the underlying mechanisms of retrieval practice on the *micro level* is beyond the scope of this article; we recommend that interested readers seek the highly informative descriptions by Carpenter et al. (2009) and Karpicke (2017).

Aftermath of the First Meeting: Two Adaptations in the Generative Learning Research Ecosphere

Inspired by the insight that retrieval practice is a potential active ingredient in tasks that had been considered as *generative tasks*, at least two types of research adaptations in generative learning have recently become apparent. First, conditions in which learners engaged in retrieval practice but not generative activities were applied more often as control conditions, in addition to or as an alternative to the usual control, namely, the restudy condition (e.g., Hoogerheide et al., 2014; Lachner et al., 2021). By this means, the potentially "hidden" retrieval-practice effect of tasks designed to elicit generative activities was supposedly offset. This research revealed beneficial effects of the tasks designed to elicit generative activities in comparison to *pure* retrieval control conditions (e.g., Endres et al., 2017; Hoogerheide et al., 2014; but see Koh et al., 2018). As retrieval was part of both tasks, these beneficial effects are likely due to the role of generative activities in enhancing the coherence and degree of elaboration of learners' mental representations. Termed differently, these findings showed that the alleged generative learning tasks are not simply *covered retrieval practice tasks*, but that generative learning activities can have added value beyond the benefits of retrieval practice.

A second adaptation of research designs in generative learning research is that the presence or absence of retrieval requirements was explicitly manipulated as a means (a) to understand the role of retrieval practice in the effects of tasks designed to elicit generative activities and (b) to optimize such tasks. Specifically, experimental comparisons were made of a *closed-book format* and *open-book format* of each task inspired by both Karpicke and Blunt (2011) and the influential study by Agarwal and colleagues (2008). The relatively few investigations having taken this path to date have yielded intriguing results (see Roelle & Nückles, 2022). What these studies demonstrated consistently is that both the quality and quantity of generative

---

[2] Note that the term elaboration is usually used differently in these two research ecospheres. For example, Carpenter et al. (2009) explain retrieval-practice effects by assuming that retrieval triggers elaboration in the sense that more associations are created between the retrieved piece of knowledge and semantically related knowledge. Such elaboration by learners is not usually considered to be intentional. The concept of elaboration in generative learning tends to be broader, entailing also the creation of novel associated knowledge presentations about, for example, analogies, example cases, or inferences. Furthermore, the associations to prior knowledge and prior experiences are emphasized as important aspects of elaboration (e.g., Nückles et al., 2020). To a large degree, learners engage in such elaboration as spontaneous and/or prompted learning strategies (e.g., spontaneous and/or prompted self-explanation).

learning activities were significantly reduced when the tasks were implemented in closed-book format. Be it the extensiveness of a concept map (see Blunt & Karpicke, 2014), the quality of self-explanations (Hiller et al., 2020), the number of generated inferences in response to adjunct questions (Roelle & Berthold, 2017), the richness of explanations in learning by teaching (e.g., Sibley et al., 2022), or the quality of written essays (Arnold et al., 2021)—the consistent finding was that compared to an open-book format, the closed-book format hindered the comprehensive execution of generative learning activities. This pattern is found in research on generative drawing as well. In studies that implemented generative drawing in a closed-book format (e.g., Kollmer et al., 2020; Schleinschok et al., 2017), the quality of learner-generated drawings appeared to be substantially lower than in studies implementing open-book generative drawing (e.g., Schmeck et al., 2014; Schwamborn et al., 2010; for a recent overview, see Fiorella & Zhang, 2018).

The obvious explanation for these effects is that retrieval is usually only fragmentary (e.g., Dunlosky & Rawson, 2015). Hence, learners assigned a closed-book task requiring generative activities often failed to retrieve all the relevant idea units from the study materials, which limited the comprehensive and accurate execution of the required generative activities (e.g., important information missing in a drawing). This detrimental effect of a closed-book format is weakened when the generative activities barely require learners to relate the respective idea units to each other (e.g., in summarization tasks). In this case, the failure to retrieve one idea unit cannot "devalue" other successfully retrieved idea units (Roelle & Berthold, 2017) and hence the detrimental effects of certain missing idea units remain relatively *local*. Nevertheless, we can conclude that a *pure* closed-book format usually hinders the comprehensive execution of the generative element in learning tasks.

In terms of learning-outcome effects, by contrast, the findings are rather mixed. Some studies reported significant effects favoring the open-book format of generative tasks (e.g., Hiller et al., 2020; Roelle & Berthold, 2017; Sibley et al., 2022), whereas other studies, just like that by Agarwal et al. (2008)[3] (which served as the main source of inspiration for this research), identified no significant differences between the two task types (e.g., Arnold et al., 2021; Roelle & Nückles, 2019; Waldeyer et al., 2020). Effects favoring a closed-book format, by contrast, are few and far between (but see Roelle & Berthold, 2017).

However, it is important to emphasize that this pattern of results should *not* be interpreted as showing that generative activities (whose execution benefits from an open-book format) would be more effective than retrieval practice or that supporting the construction of mental representations would be more important than supporting the consolidation of mental representations. Rather, the closed-book implementation of the tasks in many of these studies did not follow established recommendations on how to elicit strong retrieval-practice effects. For instance, (a) there was often practically no break between the initial learning phase and the phase(s) that included

---

[3] Note that in this study the open-book condition descriptively outperformed the closed-book condition by a medium effect size of $d = 0.45$ but the effect was not statistically significant ($p = 0.068$), potentially due to low statistical power.

retrieval practice, thus preventing substantial context updating from taking place, (b) the retrieval practice was usually not repeated, and (c) learner performance on the retrieval part was often well below 75% and some even below 51%. No reliable or substantial effects of practicing retrieval without feedback were described in Rowland's (2014) meta-analysis addressing such low retrievability of to-be-learned content (see also Karpicke, 2017). Considering the relatively low implementation quality of the tasks' retrieval practice part, the likelihood of observing beneficial retrieval practice effects was suboptimal. This aspect makes it even more interesting that in some studies the lower quality and quantity of the generative activities, which mediated detrimental effects of a closed-book format on learning outcomes (see Roelle & Nückles, 2019; Waldeyer et al., 2020), were nevertheless compensated by the higher amount of retrieval practice, resulting in null effects between the two formats concerning learning outcomes.

By contrast, studies in which learners given closed-book tasks designed to elicit generative activities had the opportunity to reinspect the learning material at some point in time reveal a different picture. In several of these studies, the closed-book format appeared to be more effective than the open-book format (e.g., Blunt & Karpicke, 2014; Waldeyer et al., 2020, Exp. 2; see also Rummer et al., 2019; but see Agarwal et al., 2008 and Wenzel et al., 2022 who detected no significant effects). It is, however, hard to attribute such benefits of the closed-book format solely to a potential consolidating function of retrieval practice, because retrieval practice has several *indirect effects* probably at work as well (see below). Nevertheless, these findings indicate that incorporating retrieval practice within tasks designed to elicit generative activities is promising. More specifically, they suggest that additionally engaging learners in retrieval practice can have added value beyond the effects of engaging learners in generative activities in terms of learning outcomes. However, this conclusion only holds if the retrieval requirement does not substantially compromise the quantity and quality of targeted generative learning activities. This condition is met when learners can review the learning material (after the initial study phase) at some time point in the learning process (see Waldeyer et al., 2020).

Although studies comparing closed- and open-book task formats designed to elicit generative activities have yielded interesting findings, they also pose many open questions. For example, it is unclear whether it is more effective to provide learners with obligatory open-book phases at fixed points in time or whether they should be given the opportunity to decide themselves when to work in a closed-book or open-book format. Also, it is unclear whether learners' skills in terms of the respective generative learning activities (e.g., their skills in generating examples or concept-mapping) are relevant and whether the formats' effectiveness depends on other learners' prerequisites. For instance, the format proposed by Waldeyer et al. (2020) in which learners could flexibly switch between a closed-book and an open-book format appears to be effective for learners with relatively low academic self-concepts (i.e., those perceiving their academic abilities and competence in a given academic domain as relatively low) but that being able to switch formats can even impair the performance more than a purely closed-book format for learners with relatively high academic self-concepts (Roelle & Renkl, 2020). Regardless of these and other open questions, it is already obvious that the inspiration originating from

retrieval practice research resulted in promising new research avenues toward generative learning deserving more research attention (see, for example, the research program recently outlined by Richter et al., 2022).

No Aftermath of the First Meeting? Few Adaptations in the Retrieval Practice Research Ecosphere

Compared to the aforementioned adaptations in generative learning research, we have the impression that retrieval practice research has been less inspired by generative learning research to date. One line of research to support this proposition is research on practice quizzing. Most practice quizzing studies seem to have mainly contextualized in retrieval practice research, emphasizing that quiz questions engage learners in practicing retrieval, thus leading to memory consolidation (direct effect of retrieval practice) and helping learners perceive their knowledge gaps (indirect effect of retrieval practice) (e.g., McDaniel et al., 2013; McDermott et al., 2014; for recent reviews, see Agarwal et al., 2021; Yang et al., 2021). However, unlike restudy or note-taking (frequent control conditions in practice quizzing research), quiz questions requiring more than just factual answers engage learners in generative learning activities such as organizing, elaborating, and generating inferences as well (e.g., Heitmann et al., 2021; Roelle et al., 2019). Hence, the effects of what is often referred to as *high- or higher-level questions* in retrieval practice research are likely driven in part by benefits of generative learning activities, which are assumed to enhance the coherence and degree of elaboration of learners' mental representations and not just by retrieval practice, which presumably mainly consolidates existing mental representations.

It is not that the benefits of such higher-level questions engaging learners in generative learning activities as well as retrieval practice are ignored in retrieval practice research. On the contrary, those questions requiring learners to venture beyond the mere retrieval of factual knowledge are considered more effective than lower-level quiz questions. However, such benefits are assumed, for instance, to stem from the stronger retrieval effort that higher-level questions demand (*desirable difficulties explanation*) or are attributed to the fact that both higher-level and lower-level quiz questions stimulate the learner to retrieve factual information; they also offer *transfer-appropriate-processing-driven* advantages when tackling higher-level questions on a final posttest (e.g., Agarwal, 2019; Jensen et al., 2014). What seems often overlooked is that quiz questions engage learners in generative activities that foster the coherence and elaboration degree of their mental representations of the content at hand. As a consequence, the functions of *constructing mental representations* (main theoretical function of generative activities) and *consolidating mental representations* (main theoretical function of retrieval practice) are seldom differentiated. In line with this focus on retrieval practice, the number of retrieved idea units or performance on given quiz questions—without differentiating reproductive from generative performance components—are frequently reported as the main learning process measures (e.g., Jensen et al., 2014; McDermott et al., 2014; see also Blunt & Karpicke, 2014).

The quantity and quality of the generated elaborations or inferences and the quality of organization reflected in learners' answers are rarely reported.

It is important to emphasize that differences in terminology between retrieval practice and generative learning research might explain parts of the alleged low degree of attendance to generative learning research on part of the retrieval practice community. For instance, studies in the retrieval practice ecosphere tend to apply the term *transfer* when referring to learners' ability to use previously acquired knowledge in novel contexts (e.g., Pan & Rickard, 2018). Hence, the term somewhat describes a learning activity and thus implies, for instance, that learners engage in inference generation (e.g., Butler, 2010). By contrast, *transfer* in generative learning research mainly describes an outcome measure, not a cognitive process. Likewise, the terms higher-order or higher-level questions in the retrieval practice community may already incorporate a description of learners' activities in response to them, and therefore, no further reference is made to the required underlying cognitive processes (e.g., organization, elaboration, or inference generation), whereas triggered cognitive activities are usually explicit in generative learning research (e.g., Roelle et al., 2019). We cannot rule out, therefore, that misunderstandings attributable to differing terminology contribute to the impression that the two ecospheres take too little note of each other's concepts, methods, and findings. This impression is reinforced by the low implementation quality of the respective *foreign* feature of learning tasks (i.e., low implementation quality of retrieval practice in generative learning research, see above, and low implementation quality of generative learning in retrieval practice research, see below).

Direct Comparisons of the Effects of Generative Learning and Retrieval Practice and How They Inspired the Investigation of Sequence Effects

The above-mentioned well-cited study of Karpicke and Blunt (2011) in which the task of open-book concept mapping (i.e., mainly generative learning, see Fig. 1) was pitted against a free-recall task (i.e., mainly retrieval practice, see Fig. 1) probably caused some dismay among the generative research community. For instance, their study pitted a task including a poorly implemented generative learning activity against a task including a well-implemented retrieval activity (as mentioned above, in the context of investigating closed- and open-book formats of tasks designed to elicit generative activities, the generative learning community made the analogous "error" and used some poorly implemented retrieval activities as well). In contrast to retrieval practice, concept mapping, like other generative activities such as journal writing, requires training and instructional support for learners to overcome utilization deficiencies (e.g., Hilbert et al., 2008; Hübner et al., 2010; Redford et al., 2012; Roelle et al., 2012), which was not provided. Also, the learning material consisted of brief expository texts with mostly factual content that hardly required learners to engage in organization, elaboration, or inference generation. Hence, it was highly improbable that concept mapping would support learning in this setting. Both

aspects might have contributed to the observation that concept mapping even failed to prove superior to repeated study of the text in Karpicke and Blunt (2011). This finding deviates substantially from the overall effect of $g = 0.58$ of concept mapping as reported in Schroeder et al.'s (2018) recent meta-analysis.

The Karpicke and Blunt study probably also caused irritation because it compared two tasks differing significantly in their main theoretical functions (constructing vs. consolidating mental representations). Such *horse race studies* (Salomon, 2002) that mainly investigate which of two or more substantially diverse learning tasks (often of unequal implementation quality) is better are widely considered suboptimal as they are not very informative (e.g., Renkl, 2015). However, such a "tricky" comparison proved quite productive in this specific case. It triggered follow-up studies that partly "repaired" the above-mentioned shortcomings associated with the implementation of concept mapping (e.g., Lechuga et al., 2015). More importantly, it stimulated in-depth analysis of the theoretical functions of the two types of learning activities: constructing coherent mental representations that are well-integrated with learner prior knowledge (generative learning activities, see Fiorella & Mayer, 2016) or consolidating existing mental representations (retrieval practice, see Karpicke, 2017).

It is fair to say that both the function of fostering coherent mental representations that are well-integrated with learner's prior knowledge and the function of fostering the consolidation of existing mental representations are important in learning. It is therefore difficult to imagine any robust arguments defending the proposal that either of these functions would be irrelevant in promoting learning. It is also likely very hard to argue whether one of these functions is more important than the other. Hence, more studies comparing learning tasks with different main functions—without taking this difference into account—probably have limited scientific value. By contrast, fruitful research in the future on how retrieval practice and generative learning research intersect should scrutinize the various functions of elicited learning activities and, on that basis, test hypotheses about *when* which type of activity is more suitable.

Sequences of Generative Learning and Retrieval Practice as a Fruitful Object of Research

When learners, after an initial study phase in which they encountered new content, are relatively far away from having understood the most important concepts and relations, engaging learners in sense-making by generative learning activities might be more beneficial than engaging them in retrieval practice. In this case, the function of supporting coherence formation and integration of new knowledge into existing knowledge structures might be more needed than consolidating (inaccurate) mental representations. By contrast, when learners already have grasped the most important content after the initial study phase, investing in the consolidation of their constructed mental representations might be the more favorable option. These assumptions were addressed in a study of Roelle and Nückles (2019), who manipulated the quality of learners' mental representations after an initial study phase

through the learning material's quality. The learners were given either a coherent and well-elaborated text on a new topic (Experiment 1) or a relatively incoherent text lacking illustrative examples (Experiment 2). The authors found that free recall was superior to a task that elicited mainly generative learning when the expository text was of high quality—in this case, engaging learners in generative activities was, similar to Karpicke and Blunt (2011), not even better than restudy. By contrast, they observed an opposite result pattern when the text was relatively incoherent and lacking illustrative examples—in this case, free recall was hardly better than restudy (though retrieval success in terms of covered idea units was similar in both experiments). The finding of O'Day and Karpicke (2021) that sequentially engaging learners in generative learning (here: open-book concept mapping) and retrieval practice revealed no added value compared to pure retrieval practice might reflect a similar cause. Potentially, the learners in O'Day and Karpicke (2021) had already understood the main concepts and relations and hence constructed a sufficiently coherent mental representation in the initial study phase, thus rendering generative activities largely redundant. The relatively poor implementation quality of concept mapping (see above, similar to the study of Karpicke and Blunt (2011)), however, might also suggest other explanations for this finding.

However, the findings and theoretical conclusions above do not imply that providing learners with tasks that mainly engage them in retrieval practice has no additional value over providing learners with tasks engaging them mainly in generative learning (i.e., open-book tasks) when they have not yet grasped the content. For example, when individual concepts rather than entirely new topics are studied in the initial learning phase, engaging learners in retrieval practice before generative learning (e.g., open-book generation of their own illustrative examples) can deliver superior learning outcomes, even when the learners initially revealed little mastery of the concepts (see Roelle et al., 2022a). One explanation for this finding is that practicing the retrieval of concept definitions can make it easier to carry out later generative activities. Likewise, poorly executed generative learning activities might even hinder subsequent retrieval. For example, when poor learner-generated illustrative examples of new concepts function as cues to retrieve concept definitions, learners may find retrieval confusing, as reflected in increased subjective extraneous load during retrieval practice (see Roelle et al., 2022a). Taking a generative-learning-before-retrieval-approach might even harm learning when learners construct flawed examples, as flawed knowledge can also be consolidated by retrieval practice (e.g., Roediger et al., 1996; Zhuang et al., 2022).

In light of the paucity of studies on different sequences of tasks eliciting mainly retrieval practice and those eliciting mainly generative activities, it is too early to draw robust conclusions about the *optimal* sequencing of such tasks. A *construction-before-consolidation* logic, that is, a generative learning-before-retrieval-practice sequence, might be more intuitive. Compared to engaging in retrieval practice before generative learning activities (open-book), higher quality mental representations would be consolidated in this case. This sequence is also better aligned with knowledge and cognitive skill acquisition theories (e.g., Bjork et al., 2013; VanLehn, 1996). These theories assume that in early phases of knowledge and skill acquisition, it would be more important to acquire basic understanding of the learning

content by constructing coherent and well-integrated mental representations before consolidating the constructed mental representations in later phases. The aforementioned empirical observations of Roelle et al., (2022a), however, seem to favor a *retrieval-practice-first sequence*, though the effect sizes were very small. There is probably *no one optimal sequence* in these two activity types. Rather, from a theoretical perspective, we can reasonably expect that there are many variables moderating the effects of *generative-first* or *retrieval-practice-first* sequences.

Potential Moderators of the Effectiveness of Sequences of Generative Learning Activities and Retrieval Practice

First, the degree to which having successfully executed one type of activity beforehand is relevant for the success of the other type of activity likely matters. When successfully performing a generative activity benefits from the learners having consolidated the required knowledge components beforehand (e.g., because learners would not need to engage in time-consuming rereading of the learning material in order to make sure that they captured all important idea units), engaging learners in well-designed retrieval practice first might pay off. If the generative activity is mastered easily without previous retrieval practice, by contrast, practicing retrieval first should not have added value compared to the opposite sequence. For instance, when learners are to generate their own examples for relatively complex new principles or concepts (e.g., Rawson & Dunlosky, 2016), practicing retrieval of their core idea units beforehand might help learners to cover all core idea units in their examples. Although learners could potentially look up the required information while generating examples, that could prove effortful and distracting, which is why previous retrieval practice might foster example quality. By contrast, if the learning content consists of very brief definitions or other *vocabulary-like* content, practicing retrieval beforehand is unlikely to improve the quality of subsequent elaborations, because learners could very quickly look up the required information (i.e., the respective *vocabulary*).

Second, aspects such as the time between the initial study phase and learners' engagement in learning activities might be relevant. When learning activities are executed immediately after studying new content, having practiced retrieval first might have relatively little effect on facilitating subsequent generative activities because the knowledge required for generative learning activities is still largely available or easily looked up, as learners might have a fairly solid grasp about which information can be accessed where in the learning material. By contrast, when learners engage in the learning tasks after a substantial delay (e.g., several days after performing tasks during a weekly seminar), making one's knowledge more readily available by practicing retrieval first might affect the quality of later generative activities substantially. The rationale behind this assumption is that when learners are not entirely sure of where to access the information they need, they might well focus on processing the information they mostly still remember. This in turn could enable elaborations that neglect certain aspects or organizational activities that overlook certain key relations between idea units.

Third, the degree to which learning tasks allow learners to take their own focusing decisions about the idea units in their learning activities probably matters as well. In journal writing (see Nückles et al., 2020), for example, learners can often decide relatively freely on which specific parts of the learning content they will focus while organizing and elaborating the learning content. In this case, engaging in retrieval practice beforehand would likely facilitate subsequent generative activities less than when generative activities are triggered by specific adjunct questions require elaborations on certain parts of the learning content. The rationale behind this assumption is that in generative tasks enabling learners to make focusing decisions on their own, learners could circumnavigate content they are unable to retrieve easily. This in turn would mean that poor retrieval would not compromise their generative output. Of course, in this case, if the content learners can recall may be minimally relevant to the learning goals, the effectiveness of generative learning activities would still be low. By contrast, when generative activities are triggered through adjunct questions, which usually focus on specific important idea units (see Endres et al., 2020) and learners cannot retrieve them, they might be impaired in generative activities regarding these idea units. In this case, practicing retrieval beforehand is more likely to benefit learners' performance on such specific learning tasks with a strong *goal-focusing component*. Notably, non-retrievable information may have detrimental effects even when the learning material is available because learners (practicing cognitive economy) may fail to try hard enough when reviewing the learning material before beginning to engage in generative activities (see Roelle et al., 2022a; Waldeyer et al., 2020).

Indirect Effects of Generative and Retrieval Practice Tasks and How They Might Complement Each Other

Retrieval practice and generative learning activities both directly influence learning outcomes and impact other factors relevant to regulative relearning. A particularly important example of *indirect effects* is how generative learning activities and retrieval practice facilitate learners' self-regulation (Endres & Renkl, 2022).

Effects on Metacognitive Monitoring and Regulation

Two central components related to current learning are *metacognitive monitoring*, whose main function is identifying acquired knowledge and knowledge gaps, and *regulation*, whose main function is closing already-identified knowledge gaps (e.g., Nelson & Narens, 1990; see also De Bruin et al., 2020). To learn effectively and efficiently, metacognitive monitoring should be as accurate as possible to (1) avoid engaging unnecessarily with already well-learned content and (2) to focus regulative (remedial) relearning on those aspects still requiring improvement (Dunlosky & Rawson, 2012). Both processes are influenced by retrieval practice and generative activities. With those tasks mainly engaging learners in retrieval practice, there is

ample evidence that metacognitive monitoring improves by working on such tasks (e.g., Carpenter et al., 2020; Hays et al., 2013; King et al., 1980; Rivers, 2021; Roediger & Karpicke, 2006). More specifically, missing or incorrect responses trigger negative feedback, either when explicit feedback originates from a learning system or teacher or more implicitly, for example, when learners become aware that they do not know the answer to a task (Endres et al., 2020; see also Weissgerber & Rummer, 2023). Working on tasks that mainly engage learners in retrieval practice thus makes learners less apt to overrate their learning success—thus resulting in more accurate monitoring, especially regarding specific elements in the learning content. Such monitoring accuracy is referred to as *metamemory*. It is important to note that these benefits are not associated with the rote learning of factual knowledge but can be found in meaningful learning as well. Several studies using texts as learning material reported improved metacognitive accuracy in retaining text segments (e.g., Endres et al., 2023; Barenberg & Dutke, 2019; Roediger & Karpicke, 2006).

Like tasks engaging learners mainly in retrieval practice, those engaging them mainly in generative activities can also foster metacognitive monitoring accuracy. However, generative learning activities mainly influence *metacomprehension*, whereas retrieval practice primarily influences metamemory. That is, engaging learners in generative activities affects how confident learners are that they have *comprehended* the learning content (e.g., Di Vesta & Finke, 1985; Griffin et al., 2019) rather than affecting their assumptions about whether a certain idea unit can be retrieved from memory. This different emphasis on metacognition's effects is probably attributable to the fact that the type of implicit cues becoming salient differs when executing generative activities or retrieval practice. The fluency and success of retrieval practice can yield valid cues enabling learners to accurately assess how well they have consolidated their memory of, for example, newly acquired declarative concepts. Such valid cues can also help them identify potential knowledge gaps. By contrast, executing generative activities provides students primarily with access to cues about the quality of their mental model regarding a problem or certain topic (see Prinz et al., 2020). Mental models, which are required to fully comprehend a problem situation, usually entail verbal and pictorial information integrated with learner's prior knowledge that is closely related to their comprehension of a given topic (Johnson-Laird, 2010; Kintsch, 1998).

The Challenges of Simultaneously Monitoring and Regulating Memory and Comprehension

As engaging in retrieval practice and generative activities improves metacognitive monitoring but steers learners' attention to various metacognitive cues, research aiming at combining these indirect effects could be another fruitful endeavor on the crossroads of generative learning and retrieval practice research. However, it will be challenging to achieve such indirect effects via tasks eliciting both activity types (e.g., closed-book tasks eliciting generative activities).

First, such tasks eliciting both activities might be metacognitively demanding because they give learners different metacognitive cues that may be partly contradictory. For example, learners might recall a central concept fluently in a task eliciting generative learning and retrieval practice, but find it very hard to generate an appropriate elaboration on that concept. In such a situation, both a *positive* retention cue and *negative* comprehension cue should ideally be considered when planning the regulation process.

Second, a challenge refers to the learners' information needs for successful regulation, that is, for closing the identified knowledge deficit. Pan and Rickard's (2018) meta-analysis suggests that in purely factual learning (e.g., pair-associative vocabulary learning), simply delivering the correct answer is sufficient. However, such simple feedback will not usually suffice during complex, meaningful learning to comprehend. More information and relearning opportunities are required in meaningful learning (e.g., elaborative feedback). In a meaningful learning context, providing only the correct answer can even detract more from successful learning than does no feedback provision at all (Pan & Rickard, 2018). Furthermore, with tasks that also elicit generative activities, it seems necessary that relearning opportunities include both elaborate instructional explanations and additional examples, as well as requiring self-explanations from the learners (Endres et al., submitted for publication). That is, more instructional support (than for the task's retrieval practice part) is needed to foster regulative relearning. Hence, both metacognitive cues during certain activities might provide contradictory information and the conditions enabling effective regulation might differ when combining the indirect effects of generative learning and retrieval practice activities.

In this context, it is an open question whether learners can effectively employ both metamemory (retrieval practice) and metacomprehension (generative learning) cues simultaneously while monitoring accurately and self-regulating effectively. There is initial evidence from Endres et al., (submitted for publication) suggesting that exploiting both cue types is feasible. The authors found that working on tasks eliciting both retrieval practice and generative activities (here: elaboration) led to more accurate monitoring than tasks focusing on retrieval practice or on generative learning only. However, the authors did not investigate whether more accurate monitoring also leads to better regulative relearning (e.g., Dunlosky et al., 2021), as learners do not always process material thoroughly enough to fully profit from this learning opportunity (e.g., Berthold & Renkl, 2010). Whether the two indirect effects can actually work together to foster learning outcomes remains unclear.

How to Maintain and Intensify the Relationship Between Research on Generative Learning and Retrieval Practice in the Future

Considering all the meeting points and intersections between retrieval practice and generative learning research described above, there are at least as many open questions as there are answers. In terms of learning tasks eliciting mainly generative activities, we have only just begun understanding which factors should be considered when deciding whether and to what degree the task should be presented in an

open-book or closed-book format and hence elicit substantial retrieval practice as well. In terms of sequentially engaging learners in mainly generative learning and retrieval practice, we presently only know that the theoretically plausible generative-first sequence is not necessarily more beneficial than a retrieval-first sequence. When combining the indirect effects of both activity types, we mainly know that such a combination is theoretically beneficial, but also potentially very challenging for learners. Keeping these complexities in mind, note that any study investigating the effects of combining generative learning and retrieval practice will naturally move our research field forward. Nevertheless, to consolidate and intensify the relationship between research on generative learning and retrieval practice, we believe that making substantial progress is subject to the three tenets below (see Fig. 2).

First, in comparing the effects of retrieval practice, generative learning activities or of them combined, it is crucial to reflect on each task's implementation quality. As mentioned above, in both research fields, the other type of learning activity has often been implemented in relatively low quality, which substantially limits what can be learned about how these two activity types interact. Hence, research focusing on combining generative learning and retrieval practice (i.e., research on open-book and closed-book task formats) should bear in mind that a single and largely unsuccessful retrieval of knowledge without feedback briefly after an initial study phase is far from optimal and hence poorly suited to analyze how these two activity types can be fruitfully combined. A similar argument relates to implementing complex generative learning activities (like concept mapping) without sufficiently introducing learners to this setting (thereby largely ignoring a plethora of learning strategy research highlighting deficits in utilization in early phases of applying a new strategy, e.g., Hübner et al., 2010; Miller, 2000). Moreover, most generative activities occur in settings involving such very simple learning material that they hardly foster learning. Such studies fail to contribute much to understanding the interplay of these two types of tasks. In terms of solving this problem, it would be more beneficial and productive if researchers from both fields were to join forces and collaborate more closely. In such collaborative studies, quality implementations of both types of activities could be realized and the implementation quality experimentally varied without triggering the suspicion that the respective "other" activity is merely serving as a straw man.

Second, even if implementation qualities are equal, it might be particularly productive to address the two types of activities as allies rather than opponents. As mentioned earlier, it is hardly plausible that one type of activity is redundant when the other type of activity is implemented as well. Rather, in view of the different theoretical main functions of generative learning activities (forming coherent mental representations well integrated with prior knowledge) and retrieval practice (consolidating acquired knowledge in memory), it appears to be more fruitful to explore how these two activity types can be combined to yield good effects on lasting meaningful learning. Studies pursuing this path could make essential contributions to the important unresolved question on how to effectively maintain the outcomes of meaningful learning, such as learners' ability to explain certain phenomena or apply knowledge in later solving problems. In rote learning of vocabulary-like content, research on *successive relearning* (e.g., Bahrick, 1979; Higham et al., 2022;

1. **Focus on implementation quality**
• Low implementation quality of retrieval practice and/or generative learning allows too little insight into the interplay of both activity types • Future studies should thus ensure high quality implementations of both activity types by adhering to established guidelines for their implementation
2. **Combination rather than competition**
• Retrieval practice and generative learning activities have so far often been treated as opponents rather than allies • Given their different main functions in theory, however, it would be most fruitful for future working groups to explore how both activity types can be effectively combined
3. **Integrating (in)direct effects**
• Studies and theoretical accounts have mostly focused solely on either direct or solely on indirect effects of retrieval practice and generative learning activities • As both effects are relevant (to grasping complex content especially), future studies should take an integrative look at both direct and indirect effects

Fig. 2 Tenets for future research on the relationship between generative learning and retrieval practice

Rawson & Dunlosky, 2011) has shown how spaced retrieval practice can contribute to maintaining factual knowledge over substantial periods of time. However, as retrieval practice appears to mainly consolidate factual knowledge (see Agarwal et al., 2021), simply having learners retrieve factual knowledge repeatedly without ensuring that learners deeply understand the respective content would not be likely to contribute to the lasting acquisition of problem-solving or explanation skills (e.g., Rawson et al., 2020). Combinations of tasks that engage learners in generative learning activities, in which learners deeply process the respective content, and tasks that engage learners mainly in retrieval practice, by contrast, are potentially useful in establishing effective knowledge acquisition and maintenance interventions in meaningful learning.

Third, up to now, most studies and even theoretical accounts have focused either on direct effects of the different activities or on their indirect effects relevant for further regulation after initial learning. Taking both effects into account is especially relevant if the learners are to understand complex content (e.g., Newton's laws or how to prove in mathematics). For such content, a simple one-presentation-plus-remediation procedure is insufficient, but the learners may have to approach such content by many iterations (in the extreme case over several school years; see the concept of a spiral curriculum; Bruner, 1960). Hence, a goal for further research would be studies and—even more importantly—theories taking an integrative perspective of direct and indirect effects.

Beyond the outlined three tenets for fruitful future research focusing on the instructional effects of varying degrees to which follow-up learning tasks engage learners in generative activities and retrieval practice, it is of course also very important to keep investigating the commonalities and differences in cognitive mechanisms that underly the benefits of generative activities and retrieval practice in future research. That is, the *educational perspective* taken in the present article would need

to be complemented by a *cognitive perspective*. This perspective could illuminate the extent to which the two activity types differ in how the assumed main benefits (forming coherent mental representations that are integrated with prior knowledge and consolidating those mental representations) are achieved and whether there are any mechanistic links between the two activity types. By showing that instructionally stimulating the two learning activities exerts promising effects and stimulating open questions in this field, the present work will hopefully inspire closer cooperation between these two research fields. Cooperative research efforts have the potential to elucidate the essential differences, similarities, and complementary effects of generative learning and retrieval practice.

Practical Implications

Finally, it is important to highlight (preliminarily) applied implications of what we know so far about (the interplay of) tasks designed to elicit mainly generative learning or mainly retrieval practice. Note that in both ecospheres, self-regulated engagement in the respective activities is investigated, as is the use of retrieval practice in exam preparation (e.g., Dunlosky et al., 2013) or the use of elaborations to deepen the understanding of learning content (Endres et al., 2021; Moning & Roelle, 2021; Nückles et al., 2020). The term *learning strategy* is standard in this case. A problem with such learning-strategy research is that (self-report) measures are very often relied upon to determine whether at all or how frequently a strategy was used, and self-reported (frequent) use of a (recommended) strategy is considered a positive indicator of strategy use (e.g., Dunn et al., 2012; Tullis & Maddox, 2020). Accordingly, students are often instructed or trained to use certain strategies more often. However, the restrictions of such an "occurrence rationale" have been outlined for years, and the respective main arguments are in line with our considerations on the effects and usefulness of retrieval practice and generative learning activities.

(1) It is not how often such strategies per se are used that counts—what does count is how well they have been used (e.g., Glogger et al., 2012; Leutner et al., 2007). As mentioned above, retrieval practice must be performed with high success rates, and it is essential that generative activities such as drawing are high quality (e.g., complete in their important elements and interrelations).

(2) The "when and why" of learning strategy use is essential (e.g., Endres et al., 2021). Hence, students need knowledge about the "when and why" of strategies, which is called conditional knowledge (e.g., Paris et al., 1983). We have emphasized the various main functions of retrieval practice and generative activities, which poses the *why* question. These different functions have also implications for the *when* question (e.g., retrieval practice when consolidation is important in a certain situation; generative learning when enhancing understanding is the sensible next step).

(3) Students' successful strategy use is often characterized by a certain profile of the strategies employed, that is, they coordinate the use of certain strategies (e.g., Glogger et al., 2012). We have discussed several aspects of how to productively (and likely less productively) coordinate retrieval practice and generative learning, either by combined tasks or by using different task sequences.

In educational practice, it would be ideal if teachers (a) used tasks eliciting (mainly) retrieval or generative activities at the right time, in appropriate form and difficulty, and in sensible combinations and (b) *additionally* informed the students about their rationale for using such tasks (i.e., the principle of informed training; Paris & Oka, 1986; see also Endres, 2023 2021). Such informed training can foster students' conditional knowledge that will help them later make informed decisions about the "when and why" of using and perhaps combining the corresponding learning activities in a self-regulated way.

References

Abel, R., & Hänze, M. (2019). Generating causal relations in scientific texts: The long-term advantages of successful generation. *Frontiers in Psychology, 10,* Article 199. https://doi.org/10.3389/fpsyg.2019.00199

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*(3), 659–701. https://doi.org/10.3102/0034654316689306

Agarwal, P. K. (2019). Retrieval practice & Bloom's taxonomy: Do students need fact knowledge before higher order learning? *Journal of Educational Psychology, 111*(2), 189–209. https://doi.org/10.1037/edu0000282

Agarwal, P. K., Karpicke, J. D., Kang, S. H. K., Roediger, H. L., III., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology, 22*(7), 861–876. https://doi.org/10.1002/acp.1391

Agarwal, P. K., Nunes, L. D., & Blunt, J. R. (2021). Retrieval practice consistently benefits student learning: A systematic review of applied research in schools and classrooms. *Educational Psychology Review, 33*(4), 1409–1453. https://doi.org/10.1007/s10648-021-09595-9

Arnold, K. M., Eliseev, E. D., Stone, A. R., McDaniel, M. A., & Marsh, E. J. (2021). Two routes to the same place: Learning from quick closed-book essays versus open-book essays. *Journal of Cognitive Psychology, 33*(3), 229–246. https://doi.org/10.1080/20445911.2021.1903011

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General, 108*(3), 296–308. https://doi.org/10.1037/0096-3445.108.3.296

Barenberg, J., & Dutke, S. (2019). Testing and metacognition: Retrieval practise effects on metacognitive monitoring in learning from text. *Memory, 27*(3), 269–279. https://doi.org/10.1080/09658211.2018.1506481

Berthold, K., & Renkl, A. (2010). How to foster active processing of explanations in instructional communication. *Educational Psychology Review, 22*(1), 25–40. https://doi.org/10.1007/s10648-010-9124-9

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: A meta-analytic review. *Memory & Cognition, 35*(2), 201–210. https://doi.org/10.3758/BF03193441

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology, 64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology, 106*(3), 849–858. https://doi.org/10.1037/a0035934

Brod, G. (2021). Generative learning: Which strategies for what age? *Educational Psychology Review, 33*(4), 1295–1318. https://doi.org/10.1007/s10648-020-09571-9

Bruner, J. S. (1960). *The process of education*. Harvard University Press.

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1118–1133. https://doi.org/10.1037/a0019902

Carpenter, S. K., Endres, T., & Hui, L. (2020). Students' use of retrieval in self-regulated learning: Implications for monitoring and regulating effortful learning experiences. *Educational Psychology Review, 32*(4), 1029–1054. https://doi.org/10.1007/s10648-020-09562-w

Carpenter, S. K., Pan, S. C., & Butler, A. C. (2022). The science of effective learning with a focus on spacing and retrieval practice. *Nature Reviews Psychology, 1*, 496–511. https://doi.org/10.1038/s44159-022-00089-1

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U.S. history facts. *Applied Cognitive Psychology, 23*(6), 760–771. https://doi.org/10.1002/acp.1507

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist, 49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

De Bruin, A. B. H., Roelle, J., Carpenter, S. K., Baars, M., & EFG-MRE,. (2020). Synthesizing cognitive load and self-regulation theory: A theoretical framework and research agenda. *Educational Psychology Review, 32*(4), 903–915. https://doi.org/10.1007/s10648-020-09576-4

Di Vesta, F. J., & Finke, F. M. (1985). Metacognition, elaboration, and knowledge acquisition: Implications for instructional design. *Educational Technology Research and Development, 33*(4), 285–293.

Dunlosky, J., Mueller, M. L., Morehead, K., Tauber, S. K., Thiede, K. W., & Metcalfe, J. (2021). Why does excellent monitoring accuracy not always produce gains in memory performance? *Zeitschrift für Psychologie, 229*(2), 104–119. https://doi.org/10.1027/2151-2604/a000441

Dunlosky, J., & Rawson, K. A. (2012). Overconfidence produces underachievement: Inaccurate self evaluations undermine students' learning and retention. *Learning and Instruction, 22*(4), 271–280. https://doi.org/10.1016/j.learninstruc.2011.08.003

Dunlosky, J., & Rawson, K. A. (2015). Practice tests, spaced practice, and successive relearning: Tips for classroom use and for guiding students' learning. *Scholarship of Teaching and Learning in Psychology, 1*(1), 72–78. https://doi.org/10.1037/stl0000024

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Dunn, K. E., Lo, W. J., Mulvenon, S. W., & Sutcliffe, R. (2012). Revisiting the motivated strategies for learning questionnaire: A theoretical and statistical reevaluation of the metacognitive self-regulation and effort regulation subscales. *Educational and Psychological Measurement, 72*(2), 312–331. https://doi.org/10.1177/0013164411413461

Endres, T. (2023). Adaptive blended learning to foster self-regulated learning – a principle-based explanation of a self-regulated learning training. In C. E. Overson, C. M. Hakala, L. L. Kordonowy, &

V. A. Benassi (Eds.), *In their own words: what scholars and teachers want you to know about why and how to apply the science of learning in your academic setting* (pp. 378–394). *Society of the Teaching of Psychology*. https://teachpsych.org/ebooks/itow

Endres, T., Carpenter, S. K., 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

Endres, T., Kranzdorf, L., Schneider, V., & Renkl, A. (2020). It matters how to recall – task differences in retrieval practice. *Instructional Science, 48*, 699–728. https://doi.org/10.1007/s11251-020-09526-1

Endres, T., Leber, J., Böttger, C., Rovers, S., & Renkl, A. (2021). Improving life-long learning by fostering students' learning strategies at university. *Psychology Learning and Teaching, 20*(1), 144–161. https://doi.org/10.1177/1475725720952025

Endres, T., & Renkl, A. (2022). Indirekte effekte von abrufübungen – intuitiv und doch häufig unterschätzt [Indirect effects of retrieval practice – intuitiv yet often underestimated]. *Unterrichtswissenschaft, 50*(2), 75–98. https://doi.org/10.1007/s42010-021-00140-9

Endres, T., Kubik V., Koslowski, K., Hahne, F., & Renkl, A. (2023). Immediate benefits of retrieval tasks: On the role of self-regulated relearning, metacognition, and motivation. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 55*(2–3), 49–66. https://doi.org/10.1026/0049-8637/a000280

Fiorella, L. (2023). Making sense of generative learning. *Educational Psychology Review*, *35*(2), Article 50 https://doi.org/10.1007/s10648-023-09769-7

Fiorella, L., & Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology, 38*(4), 281–288. https://doi.org/10.1016/j.cedpsych.2013.06.001

Fiorella, L., & Mayer, R. E. (2016). Eight ways to promote generative learning. *Educational Psychology Review, 28*(4), 717–741. https://doi.org/10.1007/s10648-015-9348-9

Fiorella, L., & Zhang, Q. (2018). Drawing boundary conditions for learning by drawing. *Educational Psychology Review, 30*(3), 1115–1137. https://doi.org/10.1007/s10648-018-9444-8

Froese, L., & Roelle, J. (2022). Expert example standards but not idea unit standards help learners accurately evaluate the quality of self-generated examples. *Metacognition and Learning, 17*(2), 565–588. https://doi.org/10.1007/s11409-022-09293-z

Froese, L., & Roelle, J. (2023). Expert example but not negative example standards help learners accurately evaluate the quality of self-generated examples. *Metacognition and Learning.* Advance online publication. https://doi.org/10.1007/s11409-023-09347-w

Glogger, I., Schwonke, R., Holzäpfel, L., Nückles, M., & Renkl, A. (2012). Learning strategies assessed by journal writing: Prediction of learning outcomes by quantity, quality, and combinations of learning strategies. *Journal of Educational Psychology, 104*(2), 452–468. https://doi.org/10.1037/a002668

Griffin, T. D., Mielicki, M. K., & Wiley, J. (2019). Improving students' metacomprehension accuracy. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 619–646). Cambridge University Press. https://doi.org/10.1017/9781108235631.025

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56*(2), 212–242. https://doi.org/10.2307/1170376

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 290–296. https://doi.org/10.1037/a0028468

Heitmann, S., Obergassel, N., Fries, S., Grund, A., Berthold, K., & Roelle, J. (2021). Adaptive practice quizzing in a university lecture: A pre-registered field experiment. *Journal of Applied Research in Memory and Cognition, 10*(4), 603–620. https://doi.org/10.1016/j.jarmac.2021.07.008

Higham, P. A., Zengel, B., Bartlett, L. K., & Hadwin, J. A. (2022). The benefits of successive relearning on multiple learning outcomes. *Journal of Educational Psychology, 114*(5), 928–944. https://doi.org/10.1037/edu0000693

Hilbert, T. S., Nückles, M., Renkl, A., Minarik, C., Reich, A., & Ruhe, K. (2008). Concept Mapping zum Lernen aus Texten: Können Prompts den Wissens- und Strategieerwerb fördern? [Concept mapping as a follow-up strategy for learning from texts: Can the acquisition of knowledge and skills be fostered by prompts?]. *Zeitschrift für Pädagogische Psychologie, 22*(2), 119–125. https://doi.org/10.1024/1010-0652.22.2.119

Hiller, S., Rumann, S., Berthold, K., & Roelle, J. (2020). Example-based learning: Should learners receive closed-book or open-book self-explanation prompts? *Instructional Science, 48*(6), 623–649. https://doi.org/10.1007/s11251-020-09523-4

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*(2), 151–164. https://doi.org/10.1016/j.jml.2013.03.002

Hoogerheide, V., Loyens, S. M. M., & van Gog, T. (2014). Effects of creating video-based modeling examples on learning and transfer. *Learning and Instruction, 33*, 108–119. https://doi.org/10.1016/j.learninstruc.2014.04.005

Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction, 20*(1), 18–29. https://doi.org/10.1016/j.learninstruc.2008.12.001

Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test … or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review, 26*(2), 307–329. https://doi.org/10.1007/s10648-013-9248-9

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceeding of the National Academy of Sciences of the United States of America (PNAS), 107*(43), 18243–18250. https://doi.org/10.1073/pnas.1012933107

Karpicke, J. D. (2017). Retrieval-based learning: a decade of progress. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (2nd ed., pp. 487–514). Academic Press. https://doi.org/10.1016/B978-0-12-809324-5.21055-9

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*(6018), 772–775. https://doi.org/10.1126/science.1199327

King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: The influence of retrieval practice. *The American Journal of Psychology, 93*(2), 329–343. https://doi.org/10.2307/1422236

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Koh, A. W. L., Lee, S. C., & Lim, S. W. H. (2018). The learning benefits of teaching: A retrieval practice hypothesis. *Applied Cognitive Psychology, 32*(3), 401–410. https://doi.org/10.1002/acp.3410

Kollmer, J., Schleinschok, K., Scheiter, K., & Eitel, A. (2020). Is drawing after learning effective for metacognitive monitoring only when supported by spatial scaffolds? *Instructional Science, 48*(5), 569–589. https://doi.org/10.1007/s11251-020-09521-6

Lachner, A., Hoogerheide, V., van Gog, T., & Renkl, A. (2022). Learning-by-teaching without audience presence or interaction: When and why does it work? *Educational Psychology Review, 34*(2), 575–607. https://doi.org/10.1007/s10648-021-09643-4

Lachner, A., Jacob, L., & Hoogerheide, V. (2021). Learning by writing explanations: is explaining to a fictitious student more effective than self-explaining? *Learning and Instruction*, *74. Article, 101438*,. https://doi.org/10.1016/j.learninstruc.2020.101438

Lechuga, M. T., Ortega-Tudela, J. M., & Gómez-Ariza, C. J. (2015). Further evidence that concept mapping is not better than repeated retrieval as a tool for learning from texts. *Learning and Instruction, 40*, 61–68. https://doi.org/10.1016/j.learninstruc.2015.08.002

Leutner, D., Leopold, C., & den Elzen-Rump, V. (2007). Self-regulated learning with a text-highlighting strategy: A training experiment. *Zeitschrift für Psychologie, 215*(3), 174–182. https://doi.org/10.1027/0044-3409.215.3.174

McDaniel, M. A. (2023). Combining retrieval practice with elaborative encoding: complementary or redundant? *Educational Psychology Review, 35. Article, 75*,. https://doi.org/10.1007/s10648-023-09784-8

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L., III. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology, 27*(3), 360–372. https://doi.org/10.1002/acp.2914

McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., III., & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied, 20*(1), 3–21. https://doi.org/10.1037/xap0000004

Miller, P. H. (2000). How best to utilize a deficiency. *Child Development, 71*(4), 1013–1017.

Moning, J., & Roelle, J. (2021). Self-regulated learning by writing learning protocols: do goal structures matter? *Learning and Instruction*, *75. Article, 101486*,. https://doi.org/10.1016/j.learninstruc.2021.101486s

Nelson, T. O., & Narens, L. (1990). Metamemory: a theoretical framework and new findings. In G. H. Bower (Ed.), *Psychology of learning and motivation* (pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Nückles, M., Roelle, J., Glogger-Frey, I., Waldeyer, J., & Renkl, A. (2020). The self-regulation-view in writing-to-learn: Using journal writing to optimize cognitive load in self-regulated learning. *Educational Psychology Review, 32*(4), 1089–1126. https://doi.org/10.1007/s10648-020-09541-1

O'Day, G. M., & Karpicke, J. D. (2021). Comparing and combining retrieval practice and concept mapping. *Journal of Educational Psychology, 113*(5), 986–997. https://doi.org/10.1037/edu0000486

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin, 144*(7), 710–756. https://doi.org/10.1037/bul0000151

Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology, 8*(3), 293–316. https://doi.org/10.1016/0361-476X(83)90018

Paris, S. G., & Oka, E. R. (1986). Self-regulated learning among exceptional children. *Exceptional Children, 53*(2), 103–108. https://doi.org/10.1177/001440298605300201

Prinz, A., Golke, S., & Wittwer, J. (2020). To what extent do situation-model-approach interventions improve relative metacomprehension accuracy? *Meta-Analytic Insights. Educational Psychology Review, 32*(4), 917–949. https://doi.org/10.1007/s10648-020-09558-6

Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General, 140*(3), 283–302. https://doi.org/10.1037/a0023956

Rawson, K. A., & Dunlosky, J. (2016). How effective is example generation for learning declarative concepts. *Educational Psychology Review, 28*(3), 649–672. https://doi.org/10.1007/s10648-016-9377-z

Rawson, K. A., Dunlosky, J., & Janes, J. L. (2020). All good things must come to an end: A potential boundary condition on the potency of successive relearning. *Educational Psychology Review, 32*(3), 851–871. https://doi.org/10.1007/s10648-020-09528-y

Redford, J. S., Thiede, K. W., Wiley, J., & Griffin, T. D. (2012). Concept mapping improves metacomprehension accuracy among 7th graders. *Learning and Instruction, 22*(4), 262–270. https://doi.org/10.1016/j.learninstruc.2011.10.007

Renkl, A. (2015). Different roads lead to Rome: the case of principle-based cognitive skills. *Learning: Research and Practice*, *1*(1), 79–90. https://doi.org/10.1080/23735082.2015.994255

Richter, T., Berger, R., Ebersbach, M., Eitel, A., Endres, T., Borromeo Ferri, R., Hänze, M., Lachner, A., Leutner, D., Lipowsky, F., Nemeth, L., Renkl, A., Roelle, J., Rummer, R., Scheiter, K., Schweppe, J., von Aufschnaiter, C., & Vorholzer, A. (2022). How to promote lasting learning in schools: Theoretical approaches and an agenda for research. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 54*(4), 135–141. https://doi.org/10.1026/0049-8637/a000258

Rivers, M. L. (2021). Metacognition about practice testing: A review of learners' beliefs, monitoring, and control of test-enhanced learning. *Educational Psychology Review, 33*(3), 823–862. https://doi.org/10.1007/s10648-020-09578-2

Roediger, H. L., III., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Roediger, H. L., III., Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: Creating false memories through repeated retrieval. *Journal of Memory and Language, 35*(2), 300–318. https://doi.org/10.1006/jmla.1996.0017

Roediger, H. L., III., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 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., & 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*(8), 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*(1), 131–147. https://doi.org/10.1037/edu0000365

Roelle, J., & Nückles, M. (2022). Der nutzen der integration von abrufübung in aufgaben zur verständnisförderung [Two in one go? The benefits of integrating retrieval practice into generative learning tasks]. *Unterrichtswissenschaft, 50*(1), 53–73. https://doi.org/10.1007/s42010-021-00134-7

Roelle, J., Froese, L., Krebs, R., Obergassel, N., & Waldeyer, J. (2022a). Sequence matters! Retrieval practice before generative learning is more effective than the reverse order. *Learning and Instruction, 80*, Article 101634. https://doi.org/10.1016/j.learninstruc.2022.101634

Roelle, J., Krüger, S., Jansen, C., & Berthold, K. (2012). The use of solved example problems for fostering strategies of self-regulated learning in journal writing. *Education Research International, Article, 751625,*. https://doi.org/10.1155/2012/751625

Roelle, J., Roelle, D., & Berthold, K. (2019). Test-based learning: Inconsistent effects between higher- and lower-level test questions. *Journal of Experimental Education, 87*(2), 299–313. https://doi.org/10.1080/00220973.2018.1434756

Roelle, J., Schweppe, J., Endres, T., Lachner, A., von Aufschnaiter, C., Renkl, A., Eitel, A., Leutner, D., Rummer, R., Scheiter, K., & Vorholzer, A. (2022b). Combining retrieval practice and generative learning in educational contexts: Promises and challenges. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 54*(4), 142–150. https://doi.org/10.1026/0049-8637/a000261

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Rummer, R., Schweppe, J., & Schwede, A. (2019). Open-book versus closed-book tests in university classes: a field experiment. *Frontiers in Psychology*, *10*, Article 463. https://doi.org/10.3389/fpsyg.2019.00463

Salomon, G. (2002). Technology and pedagogy: Why don't we see the promised revolution? *Educational Technology, 42*(2), 71–75.

Schindler, J., & Richter, T., (2023). Text generation benefits learning: A meta-analysis. *Educational Psychology Review, 35*, Article 44. https://doi.org/10.1007/s10648-023-09758-w

Schleinschok, K., Eitel, A., & Scheiter, K. (2017). Do drawing tasks improve monitoring and control during learning from text? *Learning and Instruction, 51*, 10–25. https://doi.org/10.1016/j.learninstruc.2017.02.002

Schmeck, A., Mayer, R. E., Opfermann, M., Pfeiffer, V., & Leutner, D. (2014). Drawing pictures during learning from scientific text: Testing the generative drawing effect and the prognostic drawing effect. *Contemporary Educational Psychology, 39*(4), 275–286. https://doi.org/10.1016/j.cedpsych.2014.07.003

Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and constructing concept maps: A meta-analysis. *Educational Psychology Review, 30*(2), 431–455. https://doi.org/10.1007/s10648-017-9403-9

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology, 102*(4), 872–879. https://doi.org/10.1037/a0019640

Sibley, L., Fiorella, L., & Lachner, A. (2022). It's better when I see it: Students benefit more from open-book than closed-book teaching. *Applied Cognitive Psychology, 36*(6), 1347–1355. https://doi.org/10.1002/acp.4017

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4*(6), 592–604. https://doi.org/10.1037/0278-7393.4.6.592

Tullis, J. G., & Maddox, G. B. (2020). Self-reported use of retrieval practice varies across age and domain. *Metacognition and Learning, 15*(2), 129–154. https://doi.org/10.1007/s11409-020-09223-x

VanLehn, K. (1996). Cognitive skill acquisition. *Annual Review of Psychology, 47*, 513–539. https://doi.org/10.1146/annurev.psych.47.1.513

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, 9*(3), 355–369. https://doi.org/10.1016/j.jarmac.2020.05.001

Weissgerber, S. C., & Rummer, R. (2023). More accurate than assumed: Learners' metacognitive beliefs about the effectiveness of retrieval practice. *Learning and Instruction*, *83*, Article 101678. https://doi.org/10.1016/j.learninstruc.2022.101679

Wenzel, K., Schweppe, J., & Rummer, R. (2022). Are open-book tests still as effective as closed-book tests even after a delay of two weeks? *Applied Cognitive Psychology, 36*(3), 699–707. https://doi.org/10.1002/acp.3943

Wittrock, M. C. (2010). Learning as a generative process. *Educational Psychologist, 45*(1), 40–45. https://doi.org/10.1080/00461520903433554

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin, 147*(4), 399–435. https://doi.org/10.1037/bul0000309

Zhuang, L., Wang, J., Xiong, B., Bian, C., Hao, L., Bayley, P. J., & Qin, S. (2022). Rapid neural reorganization during retrieval practice predicts subsequent long-term retention and false memory. *Nature Human Behaviour, 6*(1), 134–145. https://doi.org/10.1038/s41562-021-01188-4

Authors and Affiliations

Julian Roelle[1] · Tino Endres[2] · Roman Abel[1] · Niklas Obergassel[1] · Matthias Nückles[3] · Alexander Renkl[2]

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

[1] Faculty of Philosophy and Educational Research, Ruhr University Bochum, Bochum, Germany

[2] Department of Educational and Developmental Psychology, University of Freiburg, Freiburg, Germany

[3] Department of Educational Science, University of Freiburg, Freiburg, Germany