Learning and Instruction 49 (2017) 13–20

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Enhancing learning by retrieval: Enriching free recall with elaborative prompting

Tino Endres [a, *], Shana Carpenter [b], Alf Martin [a], Alexander Renkl [a]

[a] University of Freiburg, Department of Educational and Developmental Psychology, Engelbergerstr. 41, 79085 Freiburg, Germany
[b] Iowa State University, Department of Psychology, W112 Lagomarcino Hall, Ames, IA 50011-1041, USA

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form
20 September 2016
Accepted 22 November 2016

Keywords:
Learning by retrieval
Retrieval practice
Elaborative retrieval
Learning journals
Learning strategies
Testing effect

ABSTRACT

It is well-established in memory research that retrieval fosters learning. When applying this effect in education, it is an important question which type of retrieval task works best. Several studies have shown that learning is enhanced by linking new information with prior knowledge. A potential approach to making retrieval more effective, therefore, is to enrich retrieval instructions with the requirement to elaborate on the learning contents and link them to what is already known. In this study, we compared a free recall condition, as used in many studies on learning by retrieval, with a prompted recall condition in which learners were required to recall the information and apply it to their lives. Fifty-six undergraduate students were randomly assigned to one of these two conditions. They learned from a video-recorded lecture. One week later, learning outcomes were assessed by a posttest measuring fact recall and comprehension of the contents from the video lecture. Learners in the prompted recall group, compared to the free recall group, used more elaborative strategies in response to the recall task and achieved better comprehension scores. The effect on comprehension was mediated by the use of elaborative strategies. This pattern of results supports the constructive retrieval hypothesis, stating that retrieval is most effective when it involves constructive elaboration of the contents being learned. Our findings also encourage the use of pedagogical tasks in classroom teaching that combine elaboration and retrieval.
© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Many studies have shown that retrieval enhances learning. When students receive instruction over some information via a lecture or reading assignment, practicing to recall that information afterward increases the chances that it will be recalled again in the future. Information that students practice recalling is retained significantly better than information that they do not practice recalling (McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013; Wheeler & Roediger, 1992). Furthermore, practicing to recall information enhances later retention even when compared to alternative, non-retrieval-based strategies that involve additional exposure to the material, such as copying the information (Carpenter et al., 2016), re-reading it (Roediger & Karpicke, 2006; Roediger, Agarwal, McDaniel, & McDermott, 2011), or organizing it in a new way (Coane, 2013; Goossens, Camp, Verkoeijen, Tabbers, & Zwaan, 2014).

A large number of studies has confirmed that retrieval is an effective tool for learning a variety of information (for recent reviews, see Carpenter, 2012; Delaney, Verkoeijen, & Spirgel, 2010; Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Roediger, Putnam, & Smith, 2011; Rowland, 2014), and can be readily applied to enhancing students learning in their courses (Carpenter, Pashler, & Cepeda, 2009; Carpenter, Sachs, Martin, Schmidt, & Looft, 2012; Carpenter et al., 2016; Goossens et al., 2014; Jaeger, Eisenkraemer, & Stein, 2015; Karpicke, Blunt, Smith, & Karpicke, 2014; McDaniel, Anderson, Derbish, & Morrisette, 2007; McDaniel, Wildman, & Anderson, 2012; Roediger et al., 2011). Given these findings, practice testing has often been advocated as a pedagogical tool that should be implemented to enhance learning in educational settings (Carpenter, 2014; Pashler et al., 2007; Roediger & Pyc, 2012).

Studies exploring the benefits of retrieval typically use a fairly

* Corresponding author. Department of Psychology, University of Freiburg, Educational and Developmental Psychology, Engelbergerstr. 41, D-79085 Freiburg, Germany. Tel.: +49 (0) 761 203 9408.
E-mail addresses: tino.endres@psychologie.uni-freiburg.de (T. Endres), shacarp@iastate.edu (S. Carpenter), alf_martin@web.de (A. Martin), renkl@psychologie.uni-freiburg.de (A. Renkl).

http://dx.doi.org/10.1016/j.learninstruc.2016.11.010
0959-4752/© 2016 Elsevier Ltd. All rights reserved.

straightforward test such as free recall, cued recall, or multiple-choice, and find that retrieval is beneficial for learning. Less common, however, are efforts to modify the retrieval task in a way that produces even greater benefits on learning. Some studies have shown that higher doses of retrieval practice — that is, practicing to recall term-definitions to a criterion of three correct recalls, instead of just one — can increase the effectiveness of retrieval for both long-term retention and savings in re-learning (Rawson & Dunlosky, 2011, 2013; see also). The long-term benefits of repeated retrieval are also greater when students engage in retrieval attempts that are distributed in time, or "spaced out," relative to the same number of retrieval attempts that occur in closer proximity (e.g., Carpenter & DeLosh, 2005; Carpenter & Yeung, 2017; Pashler, Zarow, & Triplett, 2003; Pavlik & Anderson, 2005; Pyc & Rawson, 2012; Rawson & Dunlosky, 2013; Rawson, Vaughn, & Carpenter, 2015). Finally, retrieval practice can be more effective when students practice answering higher-order questions of the type that require comprehension and application, as opposed to relatively discrete fact-based questions (Jensen, McDaniel, Woodard, & Kummer, 2014).

Thus, the effectiveness of retrieval may be enhanced by spacing out repeated retrieval attempts, or by asking deep comprehension questions. Implementing these techniques in instructional settings may not be completely straightforward, however. These factors have the effect of making retrieval more challenging, which could create undesirable consequences. There is some evidence that if retrieval is too difficult, it may be ineffective, or even *counter*-effective, for some learners (e.g., see Carpenter et al., 2016; Karpicke et al., 2014; van Gog & Sweller, 2015).

An alternative approach to optimizing retrieval practice is to allow learners to retrieve the information while applying a deliberate strategy designed to maximize retention. Such a strategy can take the form of an attempt to expand upon the information being retrieved by drawing connections between concepts, or by connecting the material to prior knowledge. Studies have shown that linking new information with prior knowledge through *elaborative interrogation* — coming up with an explanation for a stated fact or concept — results in superior learning compared to simply reading the information (Pressley, McDaniel, Turnure, Wood, & Ahmad, 1987; Willoughby & Wood, 1994). Along these lines, Hinze, Wiley, and Pellegrino (2013, Experiment 2) found that, after encoding a series of scientific texts but before attempting to recall those texts, participants who were given instructions to expect an inference-based final test (e.g., involving "how" and "why" questions over processes and applications from the texts) performed better on a one-week delayed final test than those who were given instructions to expect a detail-based final test (e.g., involving "what" questions over terms and functions from the texts). In a subsequent experiment (Experiment 3) participants achieved significantly higher final test performance when they read a text followed by instructions to explain, rather than simply recall, the content from it.

Similar findings have been obtained from several studies on journal writing as a type of follow-up activity to seminar or lecture sessions (e.g., Berthold, Nückles, & Renkl, 2007; Nückles, Hübner, & Renkl, 2009; Nückles, Hübner, Dümer, & Renkl, 2010). In these studies, a learning condition quite similar to the free-recall conditions in studies of retrieval practice was used, in that learners were required to write minutes (i.e., a protocol) about what they had learned in the previous session. However, a pure free-recall type of protocol writing was usually found to be the worst option with respect to final learning outcomes. Instead, it was superior to enrich the instructions by providing prompts designed to increase such learning strategies as elaboration. In particular, prompts such as "Which examples can you think of that illustrate, confirm, or conflict with the learning contents?" fostered learning (e.g., Berthold et al., 2007; Nückles et al., 2009, 2010).

The effects of such elaborative strategies have not been systematically explored in studies of retrieval practice. Furthermore, many studies of retrieval practice have measured direct retention of relatively simple materials such as word pairs (see Rowland, 2014), so it is unknown whether the use of elaborative strategies consistently enhances retrieval-based learning of more complex materials. Indeed, a major limitation of the literature on retrieval practice is that the learning outcomes in most studies have been restricted to measures of direct retention of fairly simple materials (see Carpenter, 2012). We currently know much less about the power of retrieval to enhance the arguably more important outcomes of understanding and comprehension of complex, educationally-relevant learning contents (for discussions on this topic, see Butler, 2010; Carpenter, 2014; Pellegrino, 2012).

To that end, the current study was designed to explore the effects of an elaborative retrieval strategy on both retention and comprehension of complex, educationally-relevant material. Based on previous work showing that the learning of new information is enhanced when it can be linked with prior knowledge (Pressley et al., 1987; Willoughby & Wood, 1994), and studies of the *constructive retrieval hypothesis* (Hinze et al., 2013) demonstrating that learning is enhanced under conditions in which retrieval encourages the construction of inferences between concepts, it might be expected that retrieval-based learning is particularly effective when the retrieval conditions promote the construction of inferences based on prior knowledge.

The current study used this approach to explore learning from lectures — an educationally-realistic but seldom-used task in studies of retrieval practice (but see Butler & Roediger, 2007; Szpunar, Jing, & Schacter, 2014). After viewing a 30-min video-recorded lecture, one group of participants (the Free Recall Group) was instructed to freely recall all of the information they could remember from the video lecture, while the other group (the Prompted Recall Group) was given the same instructions supplemented with the additional instruction to provide examples from their own lives that related to the material they were recalling. One week later, both groups of participants returned for a final test containing both fact-based and comprehension-based questions from the video lecture.

Adopting a constructive retrieval view, we formulated the following hypotheses. We assumed that enriching a free-recall task with an elaborative prompt would increase the participants' use of elaborative strategies in initial recall (*Strategy Hypothesis);* testing this hypothesis can also be seen as a type of manipulation check, that is, determining whether the prompt actually elicited the intended elaboration strategies. We assumed that an elaborative prompt would enhance comprehension of the learning contents (*Learning-Outcomes Hypothesis*). We also tested the effects on fact learning. The central hypothesis of this study is that the expected effect on comprehension is mediated by the elaboration strategies employed during recall (*Mediation Hypothesis*). We aimed to test the robustness of the expected mediation effect, that is, if there were variables correlating with learning outcomes, we included them in the mediation model to clarify their role in the learning outcomes.

2. Methods

2.1. Participants

Fifty-six undergraduate students (age: $M = 23.16$, $SD = 3.42$) of different majors participated in this study. Participants were given course credit for participation. All participants were aware of taking part in research. We informed each participant about the possibility of quitting the experiment with no repercussions or disadvantage at any time. All participants provided informed consent and allowed us to use their collected data anonymously for publications.

2.2. Procedure

The experimental sessions were all computer-based. We used a three-phase procedure consisting of a learning phase (video lecture), intervention phase (Free or Prompted Recall), and an assessment phase (posttest on learning outcomes). There were two experimental sessions one week apart. In the first session participants watched the video lecture (learning phase) and then immediately worked on a free-recall task or a prompted-recall task, depending on condition. Additionally, participants were asked to rate their mental effort after the recall task (intervention phase). In the second session, we assessed learning outcomes (assessment phase).

2.3. Materials and design

We used a video-recorded lecture of 30 min about cognitive load theory (e.g., Sweller, Ayres, & Kalyuga, 2011; Sweller, van Merrienboer, & Paas, 1998, pp. 251–296). This presentation has been successfully used in a previous study on learning by journal writing (Nückles et al., 2009). We instructed the participants to watch the video lecture carefully and we informed them about the upcoming posttest.

After watching the video lecture, participants were randomly assigned to one of two conditions: Free Recall (25 participants) or Prompted Recall (30 participants). In the Free Recall Condition, we provided the following instruction: “Please, write down all contents you can remember from the just seen video lecture” (translation from German by the first author). In the Prompted Recall Condition, we enriched the free-recall instruction with an elaborative prompt: “In doing so, please refer to examples from your own life, which illustrate the learning material, are consistent with it, or stand in conflict with it”. A very similar elaboration prompt has been used successfully in earlier studies on journal writing (e.g., Glogger, Schwonke, Holzäpfel, Nückles, & Renkl, 2012; Nückles et al., 2009). Participants in both conditions had to type their answers into text boxes provided by the experimental program. To assess subjective mental effort, we asked participants directly after the (Free or Prompted) recall task to rate the mental effort they had to invest (Sweller et al., 2011). Participants indicated their mental effort using a scroll-bar (range: 1 [= low] to 9 [= high]).

The posttest was administered one week later and consisted of two sub-scales: facts and comprehension. The facts scale consisted of five questions tapping factual knowledge presented in the video lecture (e.g., “What does integrated format mean?”, see also Appendix A for the full list of items). Participants had to provide free-text answers. The internal consistency (Cronbach's alpha) of this scale was acceptable ($\alpha = 57$; see Schmitt, 1996, for a discussion of why internal consistencies over 0.5 are acceptable for learning outcome measures in meaningful learning contexts). The comprehension scale comprised four questions requiring either the application of knowledge or explanations presented in the video lecture (e.g., “In a conversation with a friend who is becoming a math teacher, she is telling you how difficult math teaching is. Today, she taught the Pythagorean Theorem. After explaining the theorem and illustrating it by an example, students should independently solve tasks. But students were not able to cope with the tasks. She did not have the feeling of successfully reaching her teaching goal. How would you explain to your friend what the problems of her teaching are from a cognitive load perspective, and how she could solve the problem?”, see Appendix B for the full list of items). Participants had to provide free-text answers. We scored the participants' answers to the posttest questions, and again obtained high interrater reliability (over both conditions, $ICC = 0.97$). The internal consistency (Cronbach's alpha) of this scale was acceptable ($\alpha = 0.71$; Schmitt, 1996).

2.4. Coding of recall protocols

We provided raters with a coding scheme for the intervention. Firstly the raters coded how many of the scientific concepts mentioned in the video lecture were included in participants' responses to the (Free or Prompted) recall task. Raters had to decide if a participant's answer contained the correct specific concept or not (e.g. integrated format, task interactivity, etc.). Twenty percent of participants' responses were coded by a second rater. We obtained high interrater reliability for determining the number of correct scientific concepts included in the protocols (over both conditions, intraclass correlation [ICC] = 0.91).

Secondly raters coded for the learning strategies used in the intervention. We provided raters with an adapted coding scheme for assessing learning-strategy indicators similar to one used in a previous study on journal writing (Glogger et al., 2012). This coding scheme referred to elaborative strategies (e.g., linking recalled information to own experiences or coming up with known examples), organizational strategies (e.g., identifying main ideas and their interrelations) and metacognitive strategies (e.g., stating problems in understanding, planning to overcome these problems). It provided a description of each strategy (i.e., elaboration, organization, and metacognition), together with prototypical examples (e.g., “Intrinsic load depends on the prior knowledge a learner has. […]. In my case a text talking about soccer is easier to read than a text about lacrosse, because I know much about this sport. Although the complexity of the topic might be similar, the intrinsic load of the soccer text should be less for me.”). For coding, raters read the recall protocols and had to decide whether a passage included a learning strategy. If a learning strategy was detected, raters had to decide which type of learning strategy was used. Twenty percent of the protocols were rated by a second rater, revealing high interrater reliability (over both conditions, $ICC = 0.91$).

2.5. Analyses

We used an alpha level of 0.05 or confidence intervals (CI) of 95% for all statistical tests. We relied on two-sided tests. Cohen's d was used as an effect size index. The values 0.20, 0.50 and 0.80 are considered as small, medium, and large effect sizes, respectively.

Data from one participant were excluded after outlier analysis. The posttest scores were more than 1.5 interquartile ranges away from those of the other participants.

3. Results

There were no significant differences between conditions (Free Recall and Prompted Recall) with respect to the number of scientific concepts from the video lecture that were included in the written responses to the recall task, $t\,(53) = 1.52$, $p = 0.135$ (for descriptive statistics see Table 1). The two groups also did not differ in the use of metacognitive strategies, $t\,(53) = 0.45$, $p = 0.655$, and organizational strategies, $t\,(51.15) = 0.13$, $p = 0.897$ (t-test for unequal variances). Furthermore, there were no group differences in the reported mental effort during the recall task, $t\,(53) = 0.097$, $p = 0.923$.

Table 1
Means and standard deviations (in brackets) of main variables in both groups.

	Free recall	Prompted recall
Intervention task		
Scientific concepts	4.96 (1.54)	4.27 (1.80)
Number of words	252.40 (103.87)	341.23 (171.59)
Time in min.	15.96 (6.97)	22.94 (10.24)
Metacognition	0.20 (0.50)	0.27 (0.58)
Organization	1.90 (0.50)	1.88 (0.66)
Elaboration	0.36 (0.64)	4.03 (3.13)
Subjective mental effort	5.72 (1.54)	5.77 (1.96)
Posttest		
Facts	5.24 (3.56)	5.26 (3.36)
Comprehension	5.56 (2.57)	7.13 (2.55)

As predicted, there was a significant difference between groups in the use of elaborative strategies, $t\ (31.86) = 6.27$, $p < 0.001$, $d = 1.62$ (t-test for unequal variances). The Prompted Recall Group used significantly more elaboration strategies than the Free Recall Group (Table 1). The Free Recall Group hardly used any elaboration strategies at all, with only 5 of the 25 participants including these strategies in their recall protocols. In contrast, 26 of the 30 participants in the Prompted Recall Group used elaboration strategies. Thus, the *Strategy Hypothesis* was confirmed.

We also checked to what extent excluding the non-compliant participants (i.e., the 4 participants not providing elaborations in the Prompted Recall Group) changed the pattern of findings. Firstly, we included them into the control group and, secondly, we excluded them from the analyses. Both procedures did not change the pattern of results.

Differences also emerged between the two groups in time-on-task and length of recall protocols (see Table 1). The Prompted Recall Group took more time than the Free Recall Group to complete the recall task, $t\ (51.13) = 2.99$, $p = 0.004$, $d = 0.80$ (t-test for unequal variances). In addition, the Prompted Recall Group wrote longer responses than the Free Recall Group, $t\ (48.73) = 2.36$, $p = 0.022$, $d = 0.63$ (t-test for unequal variances).

In summary, enriching free recall by an elaboration prompt did not affect the accuracy of recalling contents from the video lecture, perceived mental effort, or the use of organizational and metacognitive strategies. It did, however, increase the use of elaborative strategies, time-on-task, and length of responses. Thus, elaborative strategies, time spent on the recall task, and number of words produced in the recall protocols were used in our mediation analyses predicting posttest performance.

On the one-week delayed posttest, the Prompted Recall Group outperformed the Free Recall Group on comprehension questions, $t\ (53) = 2.269$, $p = 0.027$, $d = 0.62$ (Table 1), but not on fact questions, $t\ (53) = 0.029$, $p = 0.977$. This finding confirms the *Learning-Outcomes Hypothesis*. The finding that including an elaboration prompt during recall produced significant benefits on later comprehension is in line with the constructive retrieval hypothesis (Hinze et al., 2013).

We tested the *Mediation Hypothesis*, in a first step, by using the model depicted in Fig. 1. In addition to the effect of condition on the use of elaborative strategies (*Strategy Hypothesis*; path a) and the effect of condition on comprehension (*Learning-Outcomes Hypothesis*, path c), a regression analysis further showed that the use of elaborative strategies significantly predicted comprehension even when controlling for condition, $\beta = 0.298$ $p = 0.045$ (path b). A bootstrapping analysis with $m = 1000$ using Hayes process-tool (see Hayes, 2013) revealed a significant indirect effect, $\beta = 1.10$, 95% CI [0.18, 2.20] (i.e., mediation effect; see Fig. 1). The direct effect of condition on comprehension disappeared (i.e., no statistical significance) when we included elaborative strategies as a mediator ($p = 0.58$). This pattern of findings represents a full mediation (Hayes, 2013). The effect size of the mediation was medium to large, $\kappa^2 = 0.1709$ [0.0309 0.3390]; complete standardized indirect effects 0.21, CI [0.0313, 0.4286] (see Preacher & Kelley, 2011).

In a second step, we added time-on-task and number of words included in the recall protocols (both of which differed between the two conditions) as additional predictors. Using three predictors (elaborative strategies, time-on-task, and number of words), a bootstrapping analysis with $m = 1000$ using Hayes process-tool (see Hayes, 2013) revealed that time-on-task did not predict comprehension when the other two predictors were included ($\beta = 0.0008$, $p = 0.2718$, CI [-0.0007; 0.0024]). Hence, we found a model with two mediators: elaborative strategies and number of words (Fig. 2). In this model recall type (Free vs. Prompted Recall) did not have a direct influence on the number of words, but was mediated by the use of elaborative strategies. Use of elaborative strategies did not have a direct influence on comprehension, but was mediated by the number of words. The indirect effect was statistically significant, $\beta = 1.64$, 95% CI [0.67, 2.72] (i.e., mediation effect; see Fig. 2; direct effect after controlling for mediators: $p = 0.14$). This pattern of findings represented a full mediation. The effect size of the mediation indicated by the completely standardized indirect effects is 0.31, CI [0.14 0.52]. This effect size is even bigger compared to the effect size found in the first model (note that κ^2 cannot be calculated using two mediators, which is why completely standardized indirect effects are used to compare effect sizes). All other possibilities of sequencing the potential mediators were checked as well. There was no other sequencing having significant paths and indirect effects together.

Thus, the use of elaborative strategies led to longer responses which, in turn, were associated with enhanced comprehension. More fine-grained analyses revealed that the positive relationship between the number of words produced on the recall protocols and later comprehension was driven by words that reflected elaboration. A multiple regression analysis showed that the length of the recall protocols significantly predicted final comprehension scores even after controlling for the number of correct scientific concepts that were included in those protocols ($\beta = 0.45$, $t(52) = 3.40$, $p < 0.001$); the number of correct scientific concepts did not significantly predict comprehension ($\beta = 0.20$, $t(52) = 1.46$, $p = 0.151$).[1] The enhanced comprehension scores resulting from longer protocols, therefore, was not simply a product of the longer protocols being more accurate. Beyond the accuracy of initial recall, a unique portion of the variance in comprehension scores was accounted for by the degree to which participants elaborated upon the information and described how it applied to their lives. Overall, these finding confirm the *Mediation Hypothesis.*

4. Discussion

We hypothesized that enriching free recall instructions with an elaborative prompt would lead to the use of more elaborative strategies in initial recall (Strategy Hypothesis). This hypothesis was confirmed, as the Prompted Recall Group used significantly more elaborative strategies than the Free Recall Group. We further hypothesized that an elaborative prompt would enhance the comprehension of the learning contents (Learning-Outcomes Hypothesis). This hypothesis was also confirmed. The Prompted Recall

[1] Interestingly, an analog multiple regression predicting fact recall yielded a reverse pattern of results. Fact recall was significantly predicted by the recalled scientific concepts in response to the recall task ($\beta = 0.38$, $t\ (52) = 2.50$, $p < 0.016$) but not by the number of words used for elaboration ($\beta = 0.05$, $t\ (52) = -0.31$, $p = 0.754$).

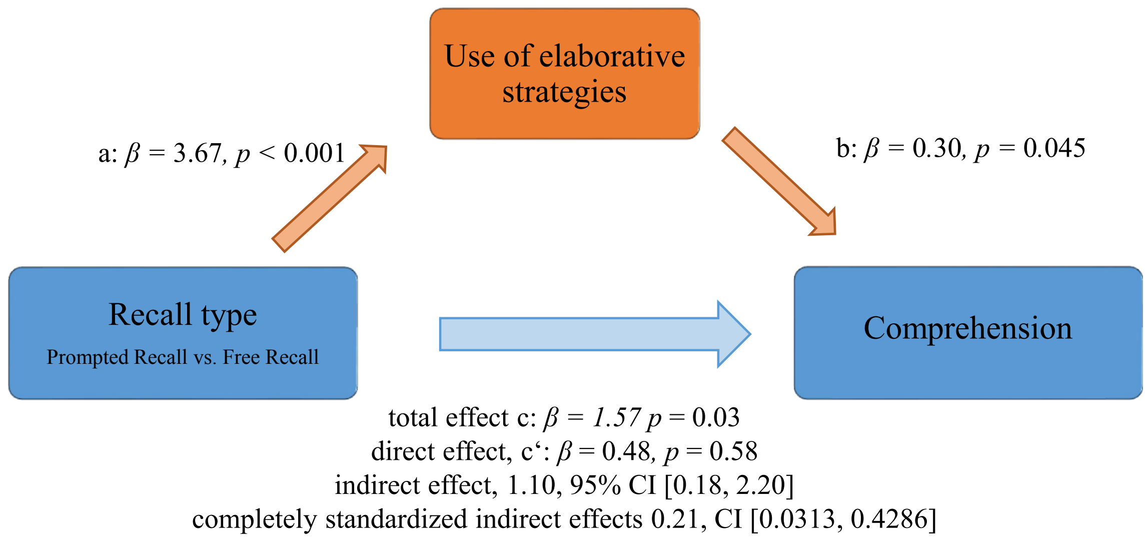

Fig. 1. Model of recall type (Prompted Recall vs. Free Recall) as a predictor of comprehension, mediated by the use of elaborative strategies. The confidence interval for the indirect effect is BCa bootstrapped CI based on 1000 samples.

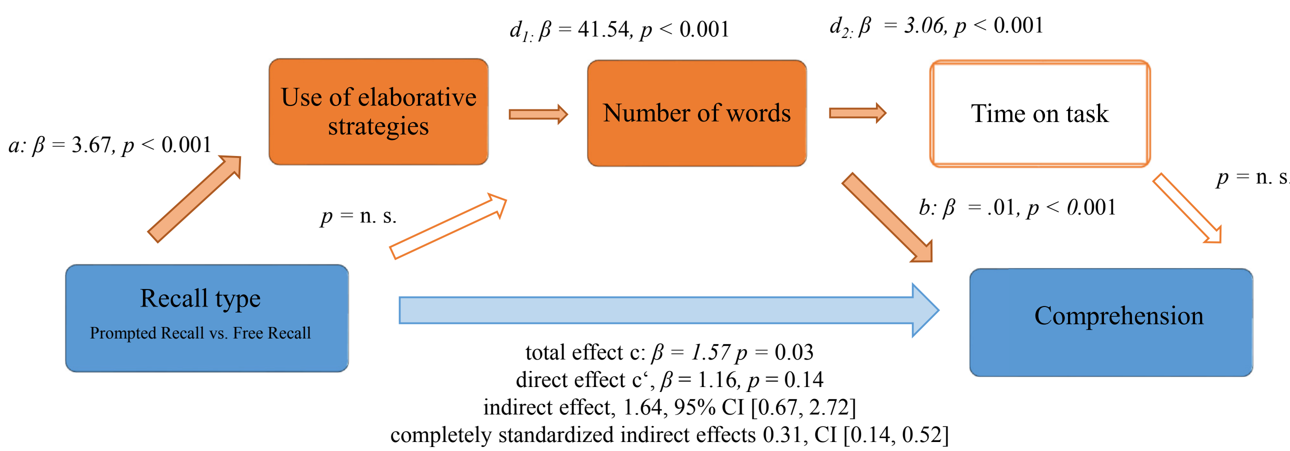

Fig. 2. Model of recall type (Prompted Recall vs. Free Recall) as a predictor of comprehension, mediated by the use of elaborative strategies and number of words. The confidence interval for the indirect effect is BCa bootstrapped CI based on 1000 samples.

Group outperformed the Free Recall Group in the comprehension of the learning materials. Fact learning did not differ between conditions. The central hypothesis of this study was that the expected effect on comprehension would be mediated by elaborative strategies employed during recall (Mediation Hypothesis*)*. This hypotheses was also confirmed.

These findings indicate that providing elaborative prompts during retrieval — that is, instructions to provide examples from one's own life illustrating the concepts recalled from a video — enhanced the effectiveness of retrieval practice on comprehension. These findings are consistent with studies demonstrating benefits of elaborative interrogation (e.g., Pressley et al., 1987), and demonstrate that a similar principle can be applied to enhance the effect of retrieval practice. More specifically, our results provide support for the *constructive retrieval hypothesis* (Hinze et al., 2013) by showing that the act of constructing connections between one's own life and the to-be-learned material during recall facilitates comprehension more than simply recalling the information without elaborative prompts.

We found that four students in the Prompted Recall Group did not provide any elaboration in their recall protocols. Excluding these students or re-assigning them to the control groups did not change our pattern of findings. There were, however, also good reasons to keep them in the Prompted Recall Group. We cannot be fully sure whether the intervention did not change the thinking patterns of the four students, although such changes were not visible in the recall protocols. It is possible that the alleged non-elaborating students thought of examples but did not right them down. Such behavior is also possible in classroom settings when our prompting intervention is applied. Hence, keeping the students in their “original” group strengthens the external validity of our findings.

Some recent studies have contributed to the on-going discussion about the educational relevance of retrieval practice (e.g., Dunlosky et al., 2013; Rawson, 2015; van Gog & Sweller, 2015). Some have explored the effects of different approaches by contrasting the learning effects of retrieval practice versus elaborative learning methods such as concept mapping (e.g., Karpicke & Blunt, 2011; Lechuga, Ortega-Tudela, & Gómez-Ariza, 2015), an imagery-based keyword method (Karpicke & Smith, 2012), or verbal elaboration (e.g., Goossens et al., 2014). The present study offers another perspective on these different methods. We did not contrast different methods, but rather combined them (i.e., combining elaborative prompting with retrieval practice). Also the findings by Blunt and Karpicke (2014. Exp. 2) showed that such combining can be productive: The effects of concept mapping were enhanced by coupling it with retrieval demands, that is, taking away the learning text whose contents should be represented in a concept map, and requiring learners to recall the information. Overall, one promising avenue for further thinking about exploiting the benefits of retrieval practice for educational purposes is to combine retrieval practice with other constructive learning methods. The elaborative prompts, as used in this study, is one example of such a combined approach.

A potential limitation of this combined approach is that it requires additional cost in words and time. The learners in the elaborative prompting condition wrote longer responses and, consequently, took more time to complete their recall task. This difference can be seen as rather “natural”, because thinking about examples and writing them down takes additional words and time. Also from a practical point of view, one might not see a major problem in the time difference. It would be fine to employ an instructional procedure (such as prompted recall) that would make students spend more productive time on follow-up course work.

On the other hand, it remains to be determined the extent to which the learning gains produced by more time-consuming methods are justified by the time and resources invested in those methods. Although our statistical findings suggest that the use of elaboration strategies and not the learning time was the crucial factor influencing comprehension, it would be theoretically and practically interesting to explore in future studies whether a similar effect on comprehension can be obtained when the amount of words and the learning time invested were the same for different methods. In doing so, we could analyse if words or time are necessary consequences of elaborative prompting or whether the effects of elaboration obtained without these costs.

Another issue for further research refers to the rating of mental effort. When the students rate their invested mental effort they might refer to rather different points of reference. Such differences in reference points might lead to a restricted comparability of the single scores. Hence, it is sensible to provide a point of reference in further studies on elaborative prompting. For example, the learners might compare their mental effort during (elaborative) recall to the effort they invested during initial learning (here: listening to a lecture).

Overall, we have shown that an elaborative prompt enhances the learning effects of retrieval practice. The use of the specific prompt was inspired by prompting effects found in studies on journal writing (e.g., Hübner, Nückles, & Renkl, 2010; Nückles et al., 2009). In those studies, sets of prompts were used, and the most effective sets included also prompts eliciting organizational and metacognitive strategies (see, e.g., Nückles et al., 2009). It is an open question whether enriching a free recall task by an organization prompt (e.g., “How can you best structure the learning contents in a meaningful way? ”) or a metacognitive prompt (“What possibilities do I have to overcome my comprehension problems?”) would have yielded similar effects on learning outcomes as the presently used elaboration prompt. One may argue that an elaborative prompt might boost retrieval practice effects, as elaborative processes have been proposed to be one of the contributors to the benefits of retrieval (Endres & Renkl, 2015; Carpenter, 2009, 2011; Rawson et al., 2015). However, there are also studies showing that retrieval practice can enhance organization of the learning contents and metacognition (e.g., Congleton & Rajaram, 2011; Roediger, Putnam, et al., 2011; Zaromb & Roediger, 2010). Hence, organizational or metacognitive prompts might further enhance additional processes responsible for retrieval practice effects. Future studies should investigate the effects of different prompts that can be used to potentially boost learning by retrieval.

In conclusion, the present study contributes new insights on retrieval-enhanced learning with complex, educationally-relevant materials. In particular, they add to a small but growing number of studies exploring the effects of retrieval on higher-level, comprehension-based outcomes (e.g., Butler, 2010; Hinze et al., 2013; Jensen et al., 2014; Larsen, Butler & Roediger, 2013). Such studies advance our understanding of retrieval-based learning by demonstrating that comprehension is significantly enhanced by engaging in constructive elaboration processes during free recall retrieval practice, and encourage further research that explores the pedagogical value of combining retrieval with activities that require learners to organize and apply the material in different ways.

Appendix. Posttest questions

A Fact based questions (translated from German by first author)

The answers to these questions could all be found in the video lecture. Hence, the participants had only to recall this information. Participants could reach from 0 to 4 points.

1 What is meant by interactivity of a task and why does this depend on the expertise of the learner?
2 What is the central executive?
3 What are schemata?
4 What is meant by “integrated format”?
5 Please explain the so-called modality effect considering the different sub-systems of human working memory.

B Comprehension questions (translated from German by the first author):

The answers to these questions were not directly provided in the lecture. Participants could reach from 0 to 4 points.

1 Empirical evidence shows that the effectiveness of worked examples decreases with increasing expertise, as compared to problem solving. That means, in later stages of skill acquisition it is sensible to have learners solve problems increasingly on their own. From a cognitive load theory perspective, this finding is called the “expertise-reversal effect”. What could this effect mean?
2 In a conversation with a friend who is becoming a math teacher, she is telling you how difficult math teaching is. Today, she taught the Pythagorean Theorem. After explaining the theorem and illustrating it by an example, students should independently solve tasks. But students were not able to cope with the tasks. She did not have the feeling of successfully reaching her teaching goal. How would you explain to your friend what the problems of her teaching are from a cognitive load perspective, and how she could solve the problem?
3 Hypermedia learning combines the idea of hypertext and multimedia approaches. Hypertext refers to a non-linear organization of information, whose reticular structure is established through logical connections (links) between knowledge units (knots). In contrast to linear texts, the sequence of processing different information is not prescribed. In the case of hypermedia learning environments, the learners additionally receive information in multimedial representations (e.g. text, photographs, figures, animations, audio, videos, etc.). The possibility of using different information formats can be seen as an advantage. However, such an advantage is accompanied by several risks. Where do you see opportunities and risks of hypermedia learning environments from the perspective of cognitive load theory?
4 Describe the three kinds of cognitive load in their relation to each other. Provide your own example which illustrates the load types in relation to each other.

References

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning & Instruction, 17*, 564–577. http://dx.doi.org/10.1016/j.learninstruc.2007.09.007.

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

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

Butler, A. C., & Roediger, H. L., III (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514–527. http://dx.doi.org/10.1080/09541440701326097.

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*, 1563–1569. http://dx.doi.org/10.1037/a0017021.

Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 37*, 1547–1552. http://dx.doi.org/10.1037/a0024140.

Carpenter, S. K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science, 21*, 279–283. http://dx.doi.org/10.1177/0963721412452728.

Carpenter, S. K. (2014). Improving student learning in low-maintenance and cost-effective ways. *Journal of Applied Research in Memory & Cognition, 3*, 121–123. http://dx.doi.org/10.1016/j.jarmac.2014.07.004.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19*, 619–636. http://dx.doi.org/10.1002/acp.1101.

Carpenter, S. K., Lund, T. J. S., Coffman, C. R., Armstrong, P. I., Lamm, M. H., & Reason, R. D. (2016). A classroom study on the relationship between student achievement and retrieval-enhanced learning. *Educational Psychology Review, 28*, 353–375. http://dx.doi.org/10.1007/s10648-015-9311-9.

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*, 760–771. http://dx.doi.org/10.1002/acp.1507.

Carpenter, S. K., Sachs, R. E., Martin, B., Schmidt, K., & Looft, R. (2012). Learning new vocabulary in German: The effects of inferring word meanings, type of feedback, and time of test. *Psychonomic Bulletin & Review, 19*, 81–86. http://dx.doi.org/10.3758/s13423-011-0185-7.

Carpenter, S. K., & Yeung, K. L. (2017). The role of mediator strength in learning from retrieval. *Journal of Memory & Language, 92*, 128–141. http://dx.doi.org/10.1016/j.jml.2016.06.008.

Coane, J. H. (2013). Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory & Cognition, 2*, 95–100. http://dx.doi.org/10.1016/j.jarmac.2013.04.001.

Congleton, A. R., & Rajaram, S. (2011). The influence of learning methods on collaboration: Prior repeated retrieval enhances retrieval organization, abolishes collaborative inhibition, and promotes post-collaborative memory. *Journal of experimental psychology: General, 140*(4), 535–551. http://dx.doi.org/10.1037/a0024308.

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 53, pp. 63–147). San Diego, CA: Elsevier. http://dx.doi.org/10.1016/S0079-7421(10)53003-2.

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*, 4–58. http://dx.doi.org/10.1177/1529100612453266.

Endres, T., & Renkl, A. (2015). Mechanisms behind the testing effect: An empirical investigation of retrieval practice in meaningful learning. *Frontiers in Psychology, 6*.

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*, 452–468. http://dx.doi.org/10.1037/a0026683.

van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: The testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review, 27*, 247–264. http://dx.doi.org/10.1007/s10648-015-9310-x.

Goossens, N. A. M. C., Camp, G., Verkoeijen, P. P. J. L., Tabbers, H. K., & Zwaan, R. A. (2014). The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning. *Journal of Applied Research in Memory & Cognition, 3*, 177–182. http://dx.doi.org/10.1016/j.jarmac.2014.05.003.

Hayes, A. (2013). *Introduction to mediation, moderation, and conditional process analysis*. New York, NY: Guilford. doi:978-1-60918-230-4.

Hinze, S. R., Wiley, J., & Pellegrino, J. W. (2013). The importance of constructive comprehension processes in learning from tests. *Journal of Memory & Language, 69*, 151–164. http://dx.doi.org/10.1016/j.jml.2013.03.002.

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

Jaeger, A., Eisenkraemer, R. E., & Stein, L. M. (2015). Test-enhanced learning in third-grade children. *Educational Psychology, 35*, 513–521. http://dx.doi.org/10.1080/01443410.2014.963030.

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*, 307–329. http://dx.doi.org/10.1007/s10648-013-9248-9.

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

Karpicke, J. D., Blunt, J. R., Smith, M. A., & Karpicke, S. S. (2014). Retrieval-based learning: The need for guided retrieval in elementary-school children. *Journal of Applied Research in Memory & Cognition, 3*, 198–206. http://dx.doi.org/10.1016/j.jarmac.2014.07.008.

Karpicke, J. D., & Smith, M. A. (2012). Separate mnemonic effects of retrieval practice and elaborative encoding. *Journal of Memory & Language, 67*, 17–29. http://dx.doi.org/10.1016/j.jml.2012.02.004.

Larsen, D. P., Butler, A. C., & Roediger, H. L., III (2013). Comparative effects of test-enhanced learning and self-explanation on long-term retention. *Medical Education, 47*, 674–682. http://dx.doi.org/10.1111/medu.12141.

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 & Instruction, 40*, 61–68. http://dx.doi.org/10.1016/j.learninstruc.2015.08.002.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*, 494–513. http://dx.doi.org/10.1080/09541440701326154.

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*, 360–372. http://dx.doi.org/10.1002/acp.2914.

McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory & Cognition, 1*, 18–26. http://dx.doi.org/10.1016/j.jarmac.2011.10.001.

Nückles, M., Hübner, S., Dümer, S., & Renkl, A. (2010). Expertise reversal effects in writing-to-learn. *Instructional Science, 38*, 237–258. http://dx.doi.org/10.1007/s11251-009-9106-9.

Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning & Instruction, 19*, 259–271. http://dx.doi.org/10.1016/j.learninstruc.2008.05.002.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., et al. (2007). *Organizing instruction and study to improve student learning (NCER 2007-2004)*. Washington, DC: U. S. Department of Education, National Center for Education Research, Institute of Education Sciences.

Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1051–1057. http://dx.doi.org/10.1037/0278-7393.29.6.1051.

Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science, 29*, 559–586. http://dx.doi.org/10.1207/s15516709cog0000_14.

Pellegrino, J. W. (2012). From cognitive principles to instructional practices: The devil is often in the details. *Journal of Applied Research in Memory & Cognition, 1*, 260–262. http://dx.doi.org/10.1016/j.jarmac.2012.10.005.

Preacher, K. J., & Kelley, K. (2011). Effect size measures for mediation models: Quantitative strategies for communicating indirect effects. *Psychological Methods, 16*, 93–115. http://dx.doi.org/10.1037/a0022658.

Pressley, M., McDaniel, M. A., Turnure, J. E., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects on intentional and incidental learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 13*, 291–300. http://dx.doi.org/10.1037/0278-7393.13.2.291.

Pyc, M. A., & Rawson, K. A. (2012). Are judgments of learning made after correct responses during retrieval practice sensitive to lag and criterion level effects? *Memory & Cognition, 40*, 976–988. http://dx.doi.org/10.3758/s13421-012-0200-x.

Rawson, K. A. (2015). The status of the testing effect for complex materials: Still a winner. *Educational Psychology Review, 27*, 327–331. http://dx.doi.org/10.1007/s10648-015-9308-4.

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*, 283–302. http://dx.doi.org/10.1037/a0023956.

Rawson, K. A., & Dunlosky, J. (2013). Relearning attenuates the benefits and costs of spacing. *Journal of Experimental Psychology: General, 142*, 1113–1129. http://dx.doi.org/10.1037/a0030498.

Rawson, K. A., Vaughn, K. E., & Carpenter, S. K. (2015). Does the benefit of testing depend on lag, an if so, why? Evaluating the elaborative retrieval hypothesis. *Memory & Cognition, 43*, 619–633. http://dx.doi.org/10.3758/s13421-014-0477-z.

Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17*, 382–395. http://dx.doi.org/10.1037/a0026252.

Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255. http://

dx.doi.org/10.1111/j.1467-9280.2006.01693.x.

Roediger, H. L., III, Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. P. Mestre, & B. H. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (Vol. 55, pp. 1–36). San Diego, CA: Elsevier. http://dx.doi.org/10.1016/b978-0-12-387691-1.00001-6.

Roediger, H. L., III, & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory & Cognition, 1*, 242–248. http://dx.doi.org/10.1016/j.jarmac.2012.09.002.

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

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*, 350–353. http://dx.doi.org/10.1037/1040-3590.8.4.350.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Heidelberg: Springer. http://dx.doi.org/10.1007/978-1-4419-8126-4.

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). *Cognitive architecture and instructional design*. Educational Psychology Review, 10. Germany: Springer. http://dx.doi.org/10.1023/A:1022193728205.

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming confidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory & Cognition, 3*, 161–164. http://dx.doi.org/10.1016/j.jarmac.2014.02.001.

Wheeler, M. A., & Roediger, H. L., III (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3*, 240–245. http://dx.doi.org/10.1111/j.1467-9280.1992.tb00036.x.

Willoughby, T., & Wood, E. (1994). Elaborative interrogation examined at encoding and retrieval. *Learning & Instruction, 4*, 139–149. http://dx.doi.org/10.1016/0959-4752(94)90008-6.

Zaromb, F. M., & Roediger, H. L. (2010). The testing effect in free recall is associated with enhanced organizational processes. *Memory & Cognition, 38*, 995–1008. http://dx.doi.org/10.3758/MC.38.8.995.