



## Would you like to learn more? Retrieval practice plus feedback can increase motivation to keep on studying

Magdalena Abel\*, Karl-Heinz T. Bäuml

*Department of Experimental Psychology, Regensburg University, Germany*



### ARTICLE INFO

**Keywords:**  
Testing effect  
Retrieval practice  
Learning  
Motivation  
Memory

### ABSTRACT

Retrieval practice can boost memory and long-term retention. The present research suggests that it may also benefit another domain that is critical for learning, namely motivation. In three experiments, subjects studied Swedish vocabulary by means of retrieval practice – with or without corrective feedback – or restudy. After a final memory test, subjects were left alone for a short period of time, and could choose freely if they wanted to learn more about the Swedish language and Sweden as a country. Retrieval practice with – but not without – feedback increased the time that subjects chose to invest in continued engagement in the materials. The results provide first evidence that retrieval practice plus feedback can increase motivation to keep on studying, potentially by making one's own progress in learning more transparent and enhancing experience of competence. Caveats and potential limitations are discussed.

### 1. Introduction

Much research examines how learners can study most efficiently. A prominent finding in this area is the so-called testing effect, demonstrating that retrieval practice during learning (“testing your memory”) boosts learning and long-term retention, usually relative to a restudy control condition (e.g., Hogan & Kintsch, 1971; Roediger III & Karpicke, 2006). This testing effect arises both in the lab and in real-world educational contexts (e.g., Greving & Richter, 2018; McDermott, Agarwal, D'Antonio, Roediger, & McDaniel, 2014). It has for instance been attributed to semantic elaboration (e.g., Carpenter, 2011) or updates of episodic context associations (e.g., Karpicke, Lehman, & Aue, 2014), with the exact mechanisms still being debated.

Retrieval practice does however not only enhance memory for practiced material, but can have further beneficial effects. Retrieval practice can benefit subsequently encoded information (e.g., Pastötter, Schicker, Niedernhuber, & Bäuml, 2011; Szpunar, McDermott, & Roediger, 2008) and additionally affect further areas relevant to efficient learning, like attention and metacognition. For instance, interpolated retrieval practice during video-recorded lectures can help students focus their attention on course contents and, by increasing engagement, enhance integration and learning of new information (e.g., Healy, Jones, Lalchandani, & Tack, 2017; Jing, Szpunar, & Schacter, 2016; Szpunar, Khan, & Schacter, 2013). Moreover, if applied during learning, retrieval practice can reduce overconfidence and allow

for more accurate metacognitive monitoring, thereby also improving self-regulated learning (e.g., Fernandez & Jamet, 2017; Karpicke, 2009; Soderstrom & Bjork, 2014; Szpunar, Jing, & Schacter, 2014).

The potential influence of retrieval practice on another key domain related to efficient learning has however not been examined yet, namely its influence on motivation. Single authors in the testing effect literature have pointed out that frequent testing and quizzing in classrooms might increase engagement by enhancing students' extrinsic motivation to study (e.g., Benjamin & Pashler, 2015; Roediger, Putnam, & Smith, 2011). The present research focused on a different question, asking if retrieval practice applied as a study strategy can also increase learners' intrinsic motivation to keep on studying – even in the absence of extrinsic reasons to do so. One of the most prominent theories in the realm of motivation science is Deci and Ryan's (1985) self-determination theory, positing that intrinsic motivation is facilitated by three factors. Besides autonomy and a sense of social relatedness, experience of competence is supposed to support the development of intrinsic motivation (see also Ryan & Deci, 2000). Importantly, this latter factor provides a potential connection between retrieval practice and intrinsic motivation.

Such a connection between retrieval practice and motivation might be expected on the basis of two reasons. The first reason is that experience of competence is fostered by optimal challenges, i.e., by tasks that are not too easy, but also not too hard to master (e.g., Danner & Lonky, 1981; Deci, 1992; Ryan & Deci, 2000). Retrieval practice may

\* Corresponding author at: Department of Experimental Psychology, Regensburg University, 93040 Regensburg, Germany.  
E-mail address: [magdalena.abel@ur.de](mailto:magdalena.abel@ur.de) (M. Abel).

provide such a challenge, creating desirable difficulties for learning by demanding greater mental effort than other study formats, like restudy (Bjork & Bjork, 2011; Pyc & Rawson, 2009). These difficulties might increase individuals' experience of competence and thus facilitate the development of intrinsic motivation. The second reason to expect a connection between retrieval practice and intrinsic motivation is that experience of competence is supported by positive feedback, making one's own progress transparent (e.g., Deci, 1971; Harackiewicz, 1979; Vallerand & Reid, 1984). Retrieval practice complemented with feedback has been shown to increase learning (e.g., Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Pashler, Cepeda, Wixted, & Rohrer, 2005) and enhance the testing effect (e.g., Kang, McDermott, & Roediger III, 2007), but it may also help learners monitor their progress in learning when applied across several practice cycles. Retrieval practice combined with feedback might therefore enhance individuals' motivation by facilitating their experience of competence in a learning task.

To examine whether retrieval practice can influence motivation, we combined the typical design of a testing-effect study with an additional phase borrowed from motivation research, the so-called free-choice period (e.g., Deci, 1971). Participants studied Swedish vocabularies and engaged in retrieval practice or restudy before completing a final test on the vocabularies. Afterwards, under a pretense, participants were left alone in the lab for a short period of time. They were free to do what they wanted during this time interval, but, among others, had the option to learn more about Sweden and the Swedish language. This free-choice period was the critical phase, and we measured how much time subjects chose to invest in further engagement in the study materials. Subjects with a higher degree of intrinsic motivation should spend more time on the materials, even in the absence of external reasons to do so.

## 2. Experiment 1

### 2.1. Method

#### 2.1.1. Ethical considerations

All reported experiments were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki. Yet, the goal was to examine the influence of learning tasks on intrinsic motivation, and if this goal had been fully disclosed before participation it would very likely have affected participants' behavior during the critical free-choice period. Therefore, before participation, subjects were only informed that the experiment examined effective learning, and that the goal was to collect data on how participants study and engage in learning. Subjects were debriefed about the full purpose of the experiment after the free-choice period was completed.

#### 2.1.2. Participants

We used previous studies on the testing effect as a starting point for determining sample size. Using this heuristic, 64 students from Regensburg University were recruited for the experiment ( $n = 32$  per condition). No psychology students were recruited, because we deemed it more likely that they might guess the full purpose of the experiment. All subjects received a small monetary reimbursement for participation. Mean age was 24.11 years ( $SD = 4.17$ ; range 19–48 years); 48 participants were female, 16 were male. One additional participant was tested, but had to be excluded prior to data analysis due to having taken Swedish classes in the past.

#### 2.1.3. Material

Material for the experimental learning task consisted of 20 Swedish-German vocabulary pairs taken from an online dictionary (<https://desv.dict.cc>). With the help of a native speaker, audio files were recorded, demonstrating the vocabulary pairs' correct pronunciations as well as those of several simple sentences in Swedish (e.g., "My name is"). All audio files were collected in a Powerpoint presentation, which also included an educational video about Sweden as a country. Materials

and data for all experiments can be found on the Open Science Framework (<https://osf.io/cwq6u/>).

#### 2.1.4. Design

The experiment had a one-factorial design, with TYPE OF PRACTICE (restudy, retrieval practice with feedback) being manipulated between participants. After initial study, subjects were asked to practice the vocabulary pairs, by means of either restudy or retrieval practice plus feedback.

#### 2.1.5. Procedure

Before each participant arrived in the lab, the experimenter activated the software HyperCam (Version 2, Hyperionics Technology, USA) on the lab computer. This software can be active in the background and monitor desktop activity, recording it as a video file without users noticing it. In particular, the software was activated to monitor activity during the free-choice period.

When subjects arrived, they were informed that the experiment examined effective learning, and that the goal was to collect data on how participants study and engage in learning. Before the memory task started, subjects were asked to provide demographic information and answer some further questions. In particular, they were asked to provide ratings on two 5-point Likert scales reflecting how interested they were in a) foreign languages more generally, and b) Swedish, specifically (1 = no interest at all; 5 = high interest). Additionally, participants were asked to indicate all foreign languages they knew, and if they had prior knowledge of Swedish.

**2.1.5.1. Memory task.** The memory task began with an initial study phase, with subjects being asked to memorize Swedish-German vocabulary pairs. All 20 vocabulary pairs were presented successively and in random order on a computer screen, at a presentation rate of 4 s per pair (e.g., eld - fire). Subsequently and without any delay, two practice cycles followed. In the retrieval practice condition, subjects were asked to try to practice the vocabulary pairs by recalling them aloud; responses were recorded by the experimenter. On each practice cycle, the 20 Swedish words were presented one at a time and for 4 s each (e.g., eld - ...?). Each trial ended with the intact Swedish-German vocabulary pair being presented as corrective feedback for another 2 s (e.g., eld - fire). When all 20 pairs had been practiced once in this manner, the second practice cycle began. In the restudy condition, subjects were on each trial presented with the intact Swedish-German vocabulary pairs for a full 6 s per pair, and were asked to try to make use of the additional study time to memorize the vocabulary as best as possible. Sequence of vocabulary pairs was random on each practice cycle and in both practice conditions.

After practice, subjects were asked to count backwards in steps of two for 60 s, and then moved on to complete a final test on all vocabulary pairs. At test, the 20 Swedish words were presented in random order and for 6 s each. Subjects were asked to recall the corresponding German words out loud, and their responses were recorded by the experimenter.

Although testing effects are often examined with longer delay between practice and test (e.g., Hogan & Kintsch, 1971; Wheeler, Ewers, & Buonanno, 2003), we decided to conduct the final test after 60 s in this experiment. On the basis of previous work (e.g., Abel & Roediger, 2018), we were hoping to achieve similar recall rates in the two practice conditions with this short delay interval, such that potential differences between the two practice conditions during the free-choice phase could not be due to different recall levels in the two conditions. Additionally, the final test was deliberately placed *before* the free-choice period because we did not want test-expectancy to act as an extrinsic motivator for engaging in the study materials during the free-choice period.

**2.1.5.2. Free-choice period.** When the final memory test was over,

participants were informed that the learning task was completed now, but that they would still have to wait in the lab for a short time. The cover story was that the experimenter had to go fetch the reimbursement money from their instructor, and that this would likely take 5 min, during which subjects should still wait in the lab. Before leaving the lab, the experimenter opened a Powerpoint presentation that subjects were free to peruse, and pointed out that it not only contained the Swedish vocabulary pairs plus audio files on their correct pronunciations, but also some simple sentences in Swedish and a video with basic information on Sweden as a country. As an alternative, the experimenter also provided the subjects with several magazines that they could go through (for similar approaches, see Deci, 1971; Harackiewicz, 1979; Murayama, Matsumoto, Izuma, & Matsumoto, 2010). The experimenter apologized for the inconvenience and told participants that they could do whatever they wanted during the next 5 min (e.g., also check their phones, or do nothing at all).

The experimenter then left, but stayed close to the lab and made sure that a constant interval of 5 min had passed before their return. This 5 min interval was the critical free-choice period. The software for monitoring desktop activity made it possible to measure if and how long subjects chose to engage in the Powerpoint presentation to learn more about Sweden and the Swedish language. To protect participants' privacy, it was ensured that the lab computer was not connected to the internet. For example, if it had been possible for subjects to check their personal email on the lab computer, private information could have been recorded inadvertently, which we made sure to avoid. After 5 min had passed, the experimenter returned to the lab. Subjects were debriefed and reimbursed for their participation.

#### 2.1.6. Analysis and coding of the free-choice period

Engagement with the Powerpoint presentation during the critical free-choice phase was coded in seconds by two independent coders; subsequently, all discrepancies in coding were resolved through discussions. Coding of engagement started when participants made cursor movements that were followed by activity involving actual contents (e.g., clicks on audio files); cursor movements alone (e.g., up and down) were not counted towards the overall engagement duration. When subjects no longer moved the cursor for 7 s or longer, this was interpreted as the end of engagement with the respective contents, and counting was stopped. For some subjects, there was only one phase of engagement with the Powerpoint presentation, equivalent to the overall duration of engagement. For others, however, there were several shorter phases of engagement, presumably interrupted by unrelated activity not involving the Powerpoint presentation. In these cases, the durations of all engagement phases were added up to form one score for overall duration of engagement with the presentation.

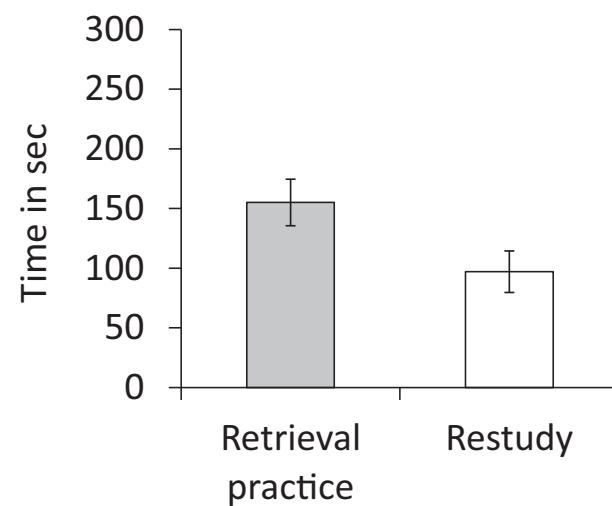
## 2.2. Results

### 2.2.1. Interest ratings

On average, subjects reported knowing 2.49 foreign languages ( $SD = 1.01$ ; range: 1–5), with no difference between practice conditions,  $t(61) < 1.00$ ,  $p = .950$ ,  $d = 0.02$ . Similarly, there were no differences between practice conditions with regard to the mean interest ratings provided on 5-point scales (1 = no interest at all; 5 = high interest); neither for mean rated interest in Swedish as a foreign language ( $M = 3.05$ ,  $SD = 1.14$ ),  $t(62) < 1.00$ ,  $p = .585$ ,  $d = 0.14$ , nor for mean rated interest in foreign languages more generally ( $M = 3.60$ ,  $SD = 1.13$ ),  $t(62) < 1.00$ ,  $p = .914$ ,  $d = 0.03$ .

### 2.2.2. Memory task

Retrieval practice plus corrective feedback enhanced performance from 43.59% correct on the first retrieval-practice cycle to 65.31% on the second retrieval-practice cycle, and to 80.78% on the final test,  $t(31) \geq 10.12$ , all  $ps \leq .001$ , all  $ds \geq 1.79$ . Final-test performance in the



**Fig. 1.** Mean time that subjects engaged in the study materials during the free-choice period in Experiment 1. Error bars represent  $\pm 1$  SEM.

restudy condition was at 74.22% and did not differ significantly from performance in the retrieval-practice condition,  $t(62) = 1.66$ ,  $p = .102$ ,  $d = 0.42$ . This finding is consistent with many studies on the testing effect, showing that the benefits of retrieval practice may often not become evident after short delay between practice and test (e.g., Abel & Roediger, 2018; Roediger & Karpicke, 2006).

### 2.2.3. Free-choice period

Fig. 1 shows the mean time (in sec) that subjects chose to keep on engaging in the Swedish Powerpoint presentation. A *t*-test for independent samples showed that subjects in the retrieval-practice condition engaged in the materials for a longer mean duration ( $M = 155.00$  sec) than subjects in the restudy condition ( $M = 96.97$  sec),  $t(62) = 2.22$ ,  $p = .030$ ,  $d = 0.56$ . Single subjects' data points were spread out across the full spectrum of 0–300 s in both conditions, but a rough breakdown of data distributions shows that subjects who kept on engaging in the material for longer time intervals were more frequently from the retrieval-practice group (see Table 1). A two-sample Kolmogorov Smirnov test confirmed that the data distributions in the two samples differed significantly,  $Z = 1.38$ ,  $p = .046$ .

## 2.3. Discussion

Retrieval practice plus feedback increased the time that subjects were willing to invest in further engagement in study materials during a free-choice period right after learning. Because recall levels and interest ratings did not differ across practice conditions, likely practice type per se induced the effect. The observed beneficial effect of retrieval practice on participants' motivation is consistent with self-determination theory (Deci & Ryan, 1985; Ryan & Deci, 2000), positing that experience of competence – as it may arise in response to effortful retrieval practice with feedback – is a critical factor for the development of intrinsic motivation.

Experiment 2 pursued two goals. The first goal was to directly

**Table 1**

Distributions of absolute numbers of subjects across different time bins for the free-choice period in Experiment 1 ( $n = 32$  for each practice condition).

Practice format	Time bin				
	0 s	1–100 s	101–200 s	201–299 s	300 s
Retrieval practice	7	4	8	7	6
Restudy	11	6	11	2	2

replicate the results of Experiment 1. The second goal was to test the limits of the beneficial effect on motivation, asking if placing a delay interval between practice and free-choice period would reduce or even eliminate the effect. For memory, retrieval practice has been shown to entail larger benefits with longer delay between practice and test, but it might well be expected that the opposite is true for the influence of retrieval practice on motivation. Even though intrinsic motivation has been associated with greater “persistence” than extrinsic motivation (e.g., Ryan & Deci, 2000), we are not aware of any work on the longevity of increases in intrinsic motivation as facilitated by an experimental learning task. The most conservative expectation, therefore, may be that motivation boosts present right after practice may fade rather quickly and not survive longer delay.

### 3. Experiment 2

#### 3.1. Method

##### 3.1.1. Participants

A power analysis for between-subjects ANOVAs indicated that a total sample of 199 subjects would be necessary to detect small- to medium-sized effects of  $f = .20$  with power at 0.80 and alpha at 0.05. 204 (non-psychology) students from Regensburg University completed the experiment. Four participants had to be excluded from data analysis due to prior knowledge of Swedish (or Norwegian, in one case). For the remaining 200 participants, mean age was 22.32 years ( $SD = 2.66$ ; range 18–33 years); 158 participants were female, 42 were male. Subjects were distributed equally across conditions ( $n = 50$  in each of four between-subjects conditions).

##### 3.1.2. Material

Material was the same as in Experiment 1.

##### 3.1.3. Design

The experiment had a  $2 \times 2$  between-subjects design. The first factor TYPE OF PRACTICE was identical to Experiment 1. Subjects were tested in one of two practice conditions, engaging in either restudy or retrieval practice plus corrective feedback after study. The second factor DELAY was new. Half of the subjects in each practice condition were asked to complete the full experiment in one session, with only a brief 1-min delay between practice and test. The other half of subjects were however asked to participate in two lab sessions. The first session ended after the practice phase, and subjects returned to the lab 2 days later for the second session to complete the final test and, subsequently, the free-choice period.

##### 3.1.4. Procedure

The experimental procedure for the short-delay condition was identical to the one applied in Experiment 1. The only difference in the long-delay condition was that the full experiment was spread out across two sessions. The first session ended after practice and after subjects had counted backwards in steps of two for 60 s. After a delay of two days, the final memory test on all vocabulary pairs was conducted at the beginning of the second session, with the subsequent free-choice period being implemented just like in the short-delay condition. Placing the final test after the delay interval imitates typical studies on the testing effect and enables a replication of the mnemonic benefits of retrieval practice.

### 3.2. Results

#### 3.2.1. Interest ratings

On average, subjects reported knowing 2.27 foreign languages ( $SD = 1.00$ ; range: 0–6), with no difference between practice conditions,  $t(198) < 1.00$ ,  $p = .357$ ,  $d = 0.13$ . There were also no differences between practice conditions regarding mean rated interest in

Swedish ( $M = 2.75$ ,  $SD = 1.02$ ),  $t(198) = 1.11$ ,  $p = .267$ ,  $d = 0.16$ , and foreign languages more generally ( $M = 3.67$ ,  $SD = 1.02$ ),  $t(198) < 1.00$ ,  $p = .678$ ,  $d = 0.06$ .

#### 3.2.2. Memory task

Retrieval practice with corrective feedback again enhanced performance from the first to the second retrieval-practice cycle (42.95% vs. 62.45%),  $t(99) = 17.07$ ,  $p < .001$ ,  $d = 1.71$ . Concerning recall on the final test, a  $2 \times 2$  ANOVA with the factors of TYPE OF PRACTICE and DELAY showed no significant main effect of TYPE OF PRACTICE,  $F(1,196) = 2.27$ ,  $MSE = 0.03$ ,  $p = .134$ ,  $\eta_p^2 = .01$ , but a significant main effect of DELAY,  $F(1,196) = 87.88$ ,  $MSE = 0.03$ ,  $p < .001$ ,  $\eta_p^2 = .31$ , reflecting reduced recall after long vs. short delay (59.55% vs. 81.95%). The ANOVA also revealed a significant interaction between the two factors,  $F(1,196) = 12.07$ ,  $MSE = 0.03$ ,  $p = .001$ ,  $\eta_p^2 = .06$ , suggesting that the role of type of practice depended on delay. Consistently, follow-up  $t$ -tests showed similar recall after retrieval practice and restudy after the 1-min delay (79.60% vs. 84.30%),  $t(98) = 1.72$ ,  $p = .089$ ,  $d = 0.34$ , but higher recall after retrieval practice relative to restudy after the 2-day interval (65.50% vs. 53.60%),  $t(98) = 3.04$ ,  $p = .003$ ,  $d = 0.61$ . This is a common pattern in the testing-effect literature, with the benefit of retrieval practice emerging after longer delay (e.g., Abel & Roediger, 2018; Roediger & Karpicke, 2006).

#### 3.2.3. Free-choice period

Fig. 2 shows the mean time (in sec) that subjects chose to keep on engaging in the Swedish Powerpoint presentation. A  $2 \times 2$  ANOVA with the factors of TYPE OF PRACTICE and DELAY showed a significant main effect of TYPE OF PRACTICE,  $F(1,196) = 9.95$ ,  $MSE = 9261.18$ ,  $p = .002$ ,  $\eta_p^2 = .05$ , but no significant main effect of DELAY,  $F(1,196) = 2.10$ ,  $MSE = 9261.18$ ,  $p = .149$ ,  $\eta_p^2 = .01$ , and also no significant interaction between the two factors,  $F(1,196) < 1.00$ ,  $MSE = 9261.18$ ,  $p = .978$ ,  $\eta_p^2 < .01$ . This suggests that practice format affected the amount of time that subjects chose to keep on engaging in the study materials, and that this influence did not vary with delay between practice and test. Follow-up  $t$ -tests confirmed that subjects decided to spend more time with the materials after retrieval practice vs. restudy, and that this was the case after both short delay ( $M = 235.88$  sec vs.  $M = 193.34$  sec),  $t(98) = 2.22$ ,  $p = .029$ ,  $d = 0.44$ , and long delay ( $M = 216.54$  sec vs.  $M = 173.24$  sec),  $t(98) = 2.24$ ,  $p = .027$ ,  $d = 0.45$ . Table 2 provides a breakdown of the data. A two-sample Kolmogorov Smirnov test showed that the data distributions in the two samples differed significantly,  $Z = 1.41$ ,  $p = .037$ .

We conducted additional analyses to address the possibility that, after long delay, subjects in the retrieval practice condition might only have spent more time on the study materials because their enhanced recall performance after delay motivated them (instead of a lasting influence of practice type per se). For this purpose, we compared the 32 subjects with the highest recall performance after longer delay in the restudy condition to 32 subjects from the long delay retrieval-practice

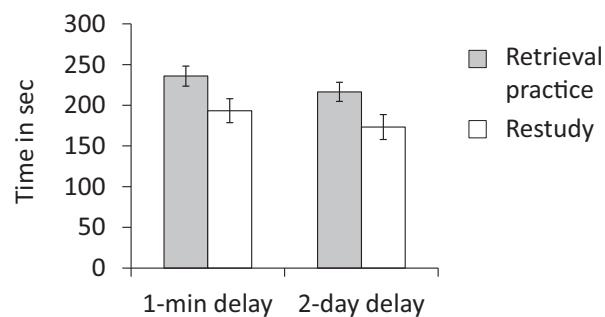


Fig. 2. Mean time that subjects engaged in the study materials during the free-choice period in Experiment 2. Error bars represent  $\pm 1$  SEM.

**Table 2**

Distributions of absolute numbers of subjects across different time bins for the free-choice period in Experiment 2 (data were collapsed across delay conditions;  $n = 100$  for each practice condition).

Practice format	Time bin				
	0 s	1–100 s	101–200 s	201–299 s	300 s
Retrieval practice	1	11	20	39	29
Restudy	8	22	19	32	19

condition that were roughly matched in recall performance (63.28% after retrieval practice vs. 65.00% after restudy). Even though final recall did not differ in these subsamples,  $t(62) < 1.00$ ,  $p = .594$ ,  $d = 0.11$ , we still found that subjects spent more time on the study materials during the free-choice period if they had previously engaged in retrieval practice plus feedback rather than restudy ( $M = 226.69$  sec vs.  $M = 163.69$  sec),  $t(62) = 2.71$ ,  $p = .009$ ,  $d = 0.68$ .

### 3.3. Discussion

The results in the short-delay condition replicate the main finding of Experiment 1 and again suggest higher motivation to keep engaging in study materials after retrieval practice plus corrective feedback than restudy. This direct replication is particularly important because Experiment 1 was the first to report any influence of retrieval practice on motivation. Recall levels in the short-delay condition of the memory task were similar in Experiments 1 and 2, but mean time investment during the free-choice period was higher in Experiment 2, indicating that different participant samples can well differ in motivation levels, but do so without influencing the effect of retrieval practice on invested time.

The results in the long-delay condition go beyond those of Experiment 1 by showing that the beneficial effect of retrieval practice on motivation can be maintained after long delay, at least with two days between practice and test. For the full collected sample, recall levels were similar after short delay, but differed between practice types after longer delay. Although this constitutes a replication of the testing effect, it may also create a confound, because the potentially persistent effect of retrieval practice plus feedback on motivation could alternatively also be attributed to the higher performance levels in the memory task, elicited right before the free-choice period. We addressed this issue in additional analyses, showing that the benefit of retrieval practice plus feedback on motivation relative to restudy was still present in subsamples of subjects that were matched in recall performance.<sup>1</sup>

According to self-determination theory (e.g., Ryan & Deci, 2000), experience of competence facilitates intrinsic motivation, and experience of competence can be enhanced by optimal challenges and feedback (e.g., Deci, 1992; Harackiewicz, 1979). Arguably, when complemented with feedback, retrieval practice could operate in both ways, providing a more challenging activity for learners than restudy and offering the opportunity to monitor progress in learning through feedback. The results of Experiments 1 and 2 are silent on which of the two factors caused the effect of retrieval practice on motivation, and we conducted Experiment 3 to address the issue. Experiment 3 was nearly identical to Experiment 1, with the main exception that restudy was contrasted with retrieval practice *without* corrective feedback. If

<sup>1</sup> To further examine this issue, future work could conduct the final memory test after short delay, but ask subjects to come back to the lab for a supposedly unrelated, second experiment after longer delay. Following this or similar approaches, it might be possible to create delay intervals without a final test directly before the free-choice period that do not simultaneously introduce a further confound with expectancy of a yet-to-be-conducted final test.

feedback is the critical component facilitating motivation in response to retrieval practice, then no benefit of retrieval practice on motivation should emerge in Experiment 3.

## 4. Experiment 3

### 4.1. Method

#### 4.1.1. Participants

Like in Experiment 1, 64 (non-psychology) students from Regensburg University participated. Mean age was 22.06 years ( $SD = 2.48$ ; range 18–30 years); 41 participants were female, 23 were male. None of the participants had any prior knowledge of Swedish.

#### 4.1.2. Material

Material was the same as in Experiment 1.

#### 4.1.3. Design

The experiment had the same one-factorial design as Experiment 1. Subjects were tested in one of two practice conditions, engaging in either restudy or retrieval practice without corrective feedback after study.

#### 4.1.4. Procedure

On retrieval-practice trials, subjects in Experiment 3 no longer received corrective feedback. To avoid recall close to floor levels in this practice phase, the Swedish words as retrieval cues were presented together with the first two initial letters of the corresponding German words. These stronger retrieval cues were presented for 6 s each, and subjects were asked to recall the full German words out loud in order to practice the vocabulary pairs. Apart from this change, all other procedural details were identical to Experiment 1.

## 4.2. Results

### 4.2.1. Interest ratings

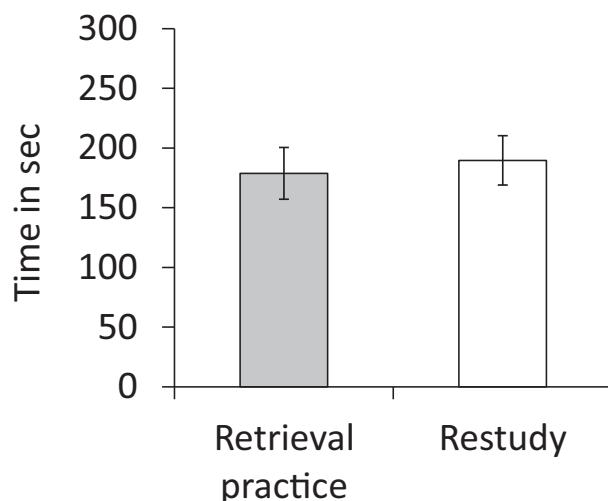
On average, subjects reported knowing 2.75 foreign languages ( $SD = 1.02$ ; range: 1–5), and there was no difference between practice conditions,  $t(62) < 1.00$ ,  $p = .333$ ,  $d = 0.24$ . In addition, there were no differences between practice conditions concerning mean rated interest in Swedish as a foreign language ( $M = 2.47$ ,  $SD = 0.96$ ),  $t(62) < 1.00$ ,  $p = .606$ ,  $d = 0.12$ , and mean rated interest in foreign languages more generally ( $M = 3.63$ ,  $SD = 1.02$ ),  $t(62) = 1.75$ ,  $p = .085$ ,  $d = 0.43$ .

### 4.2.2. Memory task

Retrieval practice without corrective feedback did not enhance performance from the first to the second retrieval-practice cycle (79.06% vs. 80.94%),  $t(31) = 1.88$ ,  $p = .070$ ,  $d = 0.33$ . Compared to the second retrieval-practice cycle, on the final test, recall in the retrieval-practice condition fell to 64.69%,  $t(31) = 7.59$ ,  $p < .001$ ,  $d = 1.34$ , caused by the fact that initial letters of German response words were provided as additional retrieval cues during practice, but not at test. In contrast, recall in the restudy condition was at 85.00% on the final test and thus superior to the retrieval-practice condition,  $t(62) = 5.71$ ,  $p < .001$ ,  $d = 1.43$ . This finding is again consistent with prior work on the testing effect. Without corrective feedback, restudy may be superior to retrieval practice, especially after shorter delay between practice and test (e.g., Kornell, Bjork, & Garcia, 2011; Wheeler et al., 2003).

### 4.2.3. Free-choice period

Fig. 3 shows the mean time (in sec) that subjects chose to keep on engaging in the Swedish Powerpoint presentation. This time, there was no difference between practice conditions, and subjects studied the materials for similar amounts of time after retrieval practice and



**Fig. 3.** Mean time that subjects engaged in the study materials during the free-choice period in Experiment 3. Error bars represent  $\pm 1$  SEM.

**Table 3**

Distributions of absolute numbers of subjects across different time bins for the free-choice period in Experiment 3 ( $n = 32$  for each practice condition).

Practice format	Time bin				
	0 s	1–100 s	101–200 s	201–299 s	300 s
Retrieval practice	6	5	6	4	11
Restudy	5	4	4	11	8

restudy ( $M = 178.75$  sec vs.  $M = 189.63$  sec),  $t(62) < 1.00$ ,  $p = .718$ ,  $d = 0.09$ . Table 3 shows a breakdown of the data, with a two-sample Kolmogorov Smirnov test suggesting that the data distributions in the two samples did not differ significantly,  $Z = 0.63$ ,  $p = .830$ .

To further evaluate this null effect, we followed Masson (2011) and applied the Bayesian information criterion (BIC) to compute posterior probabilities for the null and alternative hypotheses being correct given the observed data (D). The resulting posterior probabilities were  $P_{BIC}(H_0|D) = 0.882$  and  $P_{BIC}(H_1|D) = 0.118$ , which provides positive evidence in favor of the null hypothesis.

#### 4.3. Discussion

Retrieval practice plus feedback can affect motivation in two ways, namely by providing a more challenging activity for learners than restudy and by offering the opportunity to monitor progress in learning through feedback. Experiments 1 and 2 suggested a beneficial effect of retrieval practice on motivation, but were silent on which of the two factors caused the effect. The findings of Experiment 3, together with the results of Experiment 1, indicate that feedback critically contributes to the beneficial effect of retrieval practice on motivation. Without corrective feedback, retrieval practice no longer enhanced motivation relative to restudy. This holds while, ideally, the presence of feedback would have been manipulated in one experiment rather than between experiments.

Although the results of Experiments 1 and 3 suggest that retrieval practice does not contribute to the reported benefits for motivation – because feedback is what causes the effects – two aspects of the results deserve further clarification. First, retrieval practice in Experiment 3 was less challenging than possible, and indeed less challenging than in Experiments 1 and 2. Yet, retrieval practice was probably still a more challenging learning format than restudy, as is indicated by the fact that recall levels during practice were around 80% correct, which is much higher than one would expect if subjects were merely guessing.

Examining whether variations in difficulty of the retrieval practice task can modulate subsequent motivation is an issue of high priority for future work. Second, even if feedback caused the present effects on motivation, feedback can clearly not be provided on its own. To increase experience of competence, feedback needs to complement activities that are directed at to-be-mastered skills or contents. When the goal is to improve learning and to increase motivation to study, memory tests may constitute an optimal scaffold for feedback, because they offer a very direct and transparent impression of progress in learning, at least when applied across several practice cycles. In this manner, retrieval practice may in fact be critical for the reported benefits.

When considering the results of Experiments 1 and 2, one might like to argue that the benefits of retrieval practice plus feedback for motivation could simply have arisen because subjects noticed the repeated memory tests and expected further ones, motivating them to continue studying the materials during the free-choice period. The results of Experiment 3, however, suggest otherwise. If test expectancy was responsible for the results observed in Experiments 1 and 2, it should similarly have affected the results of Experiment 3, which is not the case. Subjects in the present experiments also seemed genuinely surprised when they were debriefed about the free-choice period having still been part of the experiments, which supports the reasoning above. Still, future work may examine the issue in more depth, for instance, by asking subjects further questions upon completion of the free-choice period and additionally probing their engagement or interest.

#### 5. General discussion

Retrieval practice with feedback during learning increased the time that subjects chose to spend on study materials later on and in the absence of extrinsic reasons to do so. This positive effect of retrieval practice relative to restudy during a free-choice period emerged with a brief delay between practice and the critical phase, but also persisted across a longer two-day delay. Retrieval practice plus feedback thus can benefit motivation to keep on studying. Retrieval practice without feedback did not affect motivation. Complementing retrieval practice with feedback thus critically contributed to the effect.

The observed beneficial effect on motivation was predicted on the basis of self-determination theory (Deci & Ryan, 1985; Ryan & Deci, 2000), which proposes that the development of intrinsic motivation is facilitated by the three factors of autonomy, social connectedness, and experience of competence. Retrieval practice can create desirable difficulties for learning by demanding greater mental effort than restudy (e.g., Pyc & Rawson, 2009), thus creating tasks that are more challenging for learners, which may increase experience of competence. Additionally, retrieval practice helps to detect gaps in one's knowledge which can then be filled on the basis of corrective feedback (e.g., Butler, Karpicke, & Roediger, 2008; Pashler et al., 2005). At least for the effect as observed in our experiments, the presence of corrective feedback seemed most critical for why retrieval practice had a positive influence on motivation. When complemented with feedback, repeated retrieval practice may provide an impression of one's progress in learning, thereby facilitating the experience of competence in mastering the task and the development of intrinsic motivation.

More research is needed to further examine the influence of learning format on motivation, but the present findings may have implications for applied settings in which learning is central. Motivation is regarded a critical component of self-regulated learning that also influences student achievement (e.g., Dweck, 1986). Based on research on the testing effect, many scientists have already started advocating for frequent testing in classrooms and the dissemination of information on retrieval practice as an effective study strategy (e.g., Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). The present experiments underline this recommendation by indicating that retrieval practice combined with feedback may not only enhance memory and long-term

retention. In addition, retrieval practice may act as a scaffold for feedback and allow students to experience competence in learning, thus potentially boosting motivation.

Indeed, in a completely separate part of the literature, educational scientists have written extensively about the role of feedback for self-regulated learning (for reviews, see Black & Wiliam, 2009; Butler & Winne, 1995), including its role for motivation (e.g., Butler, 1987; Butler & Nisan, 1986). Of particular importance for the present study, previous work in this area also examined how motivation in real classrooms is affected by formative assessment (i.e., by different methods used to evaluate student comprehension, which may also include forms of tests). For instance, Hondrich, Decristan, Hertel, and Klieme (2018) found that a structured, multi-week formative assessment treatment with personal feedback in primary school increased pupils' self-reported intrinsic motivation and perceived competence (measured by means of responses on Likert-type scales) relative to a control condition without the same structured formative assessment (for related findings, see Miller & Lavin, 2007; Pat-El, Tillema, & Van Koppen, 2012; but see, too, Yin et al., 2008). Even though studies in educational science seem to have exclusively relied on self-report measures, with no additional behavioral indicator to measure differences in motivation, these and similar findings are largely consistent with the present study. In particular, this may suggest that basic, experimental research in cognitive science can help to examine effects that have been debated in educational science.

Some caveats should be considered in future work. Experiment 1 was the first to show a benefit of retrieval practice plus feedback on motivation, and though we were able to directly replicate this effect in Experiment 2, independent replications are needed to examine the effect's reliability (e.g., Simons, 2014). Similarly, the present experiments relied on Swedish vocabulary and audio files as learning materials, and future work should examine if the benefit of retrieval practice on motivation extends to other materials and arises also outside the lab. A further issue is that, ideally, experimenters should be blinded to practice condition so that demand characteristics cannot impact the results. Experimenters were not blinded in the present study, and figuring out how to do this while still experimentally manipulating practice type will pose a challenge to future work. Last but not least, the present research is silent on whether subjects actually learnt anything during the free-choice period. Subjects chose to spend more time on going through the materials after retrieval practice, but as is, we have no way of knowing if they actually picked up any new information during this interval or just browsed through the materials rather mindlessly. Demonstrating that the increased motivation observed in this study comes with additional learning would strongly add to the practical relevance of the present findings.

## 6. Conclusions

This manuscript reported a novel finding on how retrieval practice plus corrective feedback may benefit learning. Most previous work on retrieval practice examined its benefits for memory and long-term retention, but other recent work has indicated that retrieval practice may additionally also boost metacognition and attention. Here, we showed that retrieval practice with feedback may also benefit intrinsic motivation to keep on studying, which is regarded an important factor for self-regulated learning. The finding fits well within the larger testing effect literature and adds to the overall picture that retrieval practice is generally beneficial for learning. In addition, we hope that the finding may stimulate further experimental research on the precise role of learning format for motivation, an issue that seems to have been largely neglected by cognitive scientists to date.

## Authors' note

This research did not receive any specific grant from funding

agencies in the public, commercial, or not-for-profit sectors. The data and materials for all experiments are available on the Open Science Framework (<https://osf.io/cwq6u/>).

## CRediT authorship contribution statement

**Magdalena Abel:**Conceptualization, Methodology, Formal analysis, Writing - original draft. **Karl-Heinz T. Bäuml:**Conceptualization, Methodology, Writing - review & editing.

## Declaration of competing interest

The authors have no competing interests to declare.

## References

- Abel, M., & Roediger, H. L. (2018). The testing effect in a social setting: Does retrieval practice benefit a listener? *Journal of Experimental Psychology: Applied*, 24, 347–359.
- Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213–238.
- Benjamin, A. S., & Pashler, H. (2015). The value of standardized testing: A perspective from cognitive psychology. *Policy Insights From the Behavioral and Brain Sciences*, 2, 13–23.
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.). *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56–64). New York: Worth Publishers.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21, 5–31.
- Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 918–928.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65, 245–281.
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79, 474–482.
- Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology*, 78, 210–216.
- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1547–1552.
- Danner, F. W., & Lonky, E. (1981). A cognitive-developmental approach to the effects of rewards on intrinsic motivation. *Child Development*, 52, 1043–1052.
- Deci, E. L. (1971). Effects on externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 18, 105–115.
- Deci, E. L. (1992). The relation of interest to the motivation of behavior: A self-determination theory perspective. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.). *The role of interest in learning and development* (pp. 43–70). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York, NY: Plenum Press.
- 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.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040–1048.
- Fernandez, J., & Jamet, E. (2017). Extending the testing effect to self-regulated learning. *Metacognition and Learning*, 12, 131–156.
- Greving, S., & Richter, T. (2018). Examining the testing effect in university teaching: Retrievalability and question format matter. *Frontiers in Psychology*, 9, 2412.
- Harackiewicz, J. M. (1979). The effects of reward contingency and performance feedback on intrinsic motivation. *Journal of Personality and Social Psychology*, 37, 1352–1363.
- Healy, A. F., Jones, M., Lalchandani, L. A., & Tack, L. A. (2017). Timing of quizzes during learning: Effects of motivation and retention. *Journal of Experimental Psychology: Applied*, 23, 128–137.
- Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 10, 562–567.
- Hondrich, A. L., Decristan, J., Hertel, S., & Klieme, E. (2018). Formative assessment and intrinsic motivation: The mediating role of perceived competence. *Zeitschrift für Erziehungswissenschaft*, 21, 717–734.
- Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied*, 22, 305–318.
- Kang, S. H., McDermott, K. B., & Roediger, H. L., III (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of*

- Cognitive Psychology*, 19, 528–558.
- Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, 138, 469–486.
- Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In B. H. Ross (Vol. Ed.), *Psychology of learning and motivation*. Vol. 61. *Psychology of learning and motivation* (pp. 237–284). San Diego: Elsevier Academic Press.
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, 65, 85–97.
- Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, 43, 679–690.
- McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., & 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, 3–21.
- Miller, D., & Lavin, F. (2007). “But now I feel I want to give it a try”: Formative assessment, self-esteem and a sense of competence. *The Curriculum Journal*, 18, 3–25.
- Murayama, K., Matsumoto, M., Izuma, K., & Matsumoto, K. (2010). Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proceedings of the National Academy of Sciences*, 107, 20911–20916.
- Pashler, H., Cepeda, N. J., Wixted, J., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 3–8.
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 37, 287–297.
- Pat-El, R., Tillemans, H., & Van Koppen, S. W. M. (2012). Effects of formative feedback on intrinsic motivation: Examining ethnic differences. *Learning and Individual Differences*, 22, 449–454.
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60, 437–447.
- Roediger, H. L., Putnam, A. L., & Smith, M. A. (2011). Ten benefits of testing and their applications to educational practice. In J. Mestre, & B. Ross (Eds.). *Psychology of learning and motivation: Cognition in education* (pp. 1–36). Oxford: Elsevier.
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255.
- Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78.
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9, 76–80.
- Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language*, 73, 99–115.
- Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, 3, 161–164.
- Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, 110, 6313–6317.
- Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1392–1399.
- Vallerand, R. J., & Reid, G. (1984). On the causal effects of perceived competence on intrinsic motivation: A test of cognitive evaluation theory. *Journal of Sport and Exercise Psychology*, 6, 94–102.
- Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, 11, 571–580.
- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., & Furtak, E. M. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education*, 21, 335–359.