When Study Capacities are Limited and Deadline is Fixed – How Practice Type and Practice Timing Influence Recall of Practiced and Unpracticed Material

Lisa Wallner, Anna T. Nickl, & Karl-Heinz T. Bäuml*

Department of Experimental Psychology, Regensburg University, Germany

To appear in:

Journal of Applied Research in Memory and Cognition

**Correspondence Address:*

Karl-Heinz T. Bäuml

Department of Experimental Psychology, Regensburg University,

93040 Regensburg, Germany.

E-mail: karl-heinz.baeuml@ur.de

Phone: +49-941-943-3818

FAX: +49-941-943-3872

Abstract

Addressing a common problem for students – limited study capacities and a fixed deadline – this study investigated how and when selective practice should be implemented to maximize recall of practiced and unpracticed items. In each of three experiments, participants studied a list of items and recalled the material 48 hours later. Immediately after study, 24 hours after study, or shortly before test, participants practiced some of the items through restudy (Experiment 1), retrieval practice without feedback (Experiment 2), or retrieval practice plus feedback (Experiment 3). Results showed that, on average, recall was best after retrieval practice plus feedback and worst after retrieval practice without feedback, and it improved from early to late practice. Consistently, recall was highest after late retrieval practice plus feedback. Results also point to pronounced differences between practice types regarding the effects of practice timing and the relative contributions of practiced and unpracticed material to test performance.

Keywords

restudy – retrieval – feedback – practice type – practice timing

General Audience Summary

A common problem for students when preparing for a test that is set to take place at a fixed time in the future is knowing that they will not be able to study every single detail of the subject matter at hand – due to other tasks, a job, or family obligations. In such situations, it is imperative to know when and how they should practice the study material in order to remember as much as possible at test – of both the materials they got to practice and those they did not have time for. To address this question, we conducted three experiments, in each of which we presented participants with a list of words and later provided them with an opportunity to practice some of the words. Two days after study, the participants were tested on all studied words. We varied timing and type of practice: participants practiced either directly after study, one day later, or directly before the final test; and they practiced through restudy, i.e., rereading and memorizing, practicing retrieval, or practicing retrieval and then receiving the correct responses as feedback. The results showed that, on average, (i) final recall was best after practicing retrieval and receiving feedback, and worst after practicing retrieval without receiving feedback, and (ii) final recall improved from early to late practice. Most important, final recall was highest after late retrieval practice with feedback and this effect was largely driven by high recall levels for the practiced items. These findings suggest that students should practice study material through practicing retrieval and providing themselves with feedback, and that they should do this shortly before taking the final test. Future research may examine generalizability of the results, for instance, by examining whether the results extend to more complex study material and prolonged retention intervals.

When Study Capacities are Limited and Deadline is Fixed – How Practice Type and Practice Timing Influence Recall of Practiced and Unpracticed Material

Imagine a student who attends a lecture and knows she will be tested on its subject matter two days later. Pressed for time, she fears she will be able to practice only parts of the material, although her definitive goal is to remember as much as possible of the lecture contents – both of the practiced and the unpracticed contents. Aiming to maximize her final test score, she asks two questions: First, *how* should she practice the material – by reading through it again (restudy), quizzing herself on the material (retrieval practice), or quizzing herself and then looking up the correct answers (retrieval practice plus feedback)? And second, *when* should she practice the material – immediately after the lecture, the next day (and thus one day before the test), or directly before the test?

In-depth results on the effects of different combinations of practice type (restudy versus retrieval practice versus retrieval practice plus feedback) and practice timing (early versus middle versus late practice during a constant longer retention interval) on final test performance in the literature are scarce. Indeed, to the best of our knowledge, there is no single study yet that manipulated both type and timing of selective practice while keeping retention interval between study and test constant. This holds although some studies provide at least fragmentary information on the issue. For instance, several studies examined the effects of one or two types of selective practice, with practice taking place either early or late during a longer retention interval. A number of other studies examined the effects of one or two types of nonselective practice, with practice taking place early during a longer retention interval. Also, some studies compared the effects of nonselective retrieval practice and restudy on final recall for two practice times, shortly after study and later during a longer retention interval. The following paragraphs shortly summarize this literature.

Studies on Selective Practice

In studies on the effects of selective practice, only a subset of the studied material is practiced after study before, on a later final test, recall of all studied items is measured – both of the practiced and the unpracticed material. When such selective practice occurs immediately after study and re-call is tested shortly after practice, selective retrieval practice without feedback typically induces a beneficial effect for the practiced material and a detrimental effect for the unpracticed material, relative to an appropriate control condition (Anderson et al., 1994; Anderson & Spellman, 1995). This basic finding of beneficial and detrimental effects generalizes to the case when practice occurs through retrieval practice plus feedback (Ciranni & Shimamura, 1999; Erdman & Chan, 2013), while only the beneficial effect for the practiced material may be found after selective restudy (Anderson et al., 2000; Bäuml & Aslan, 2004; for details, see Bäuml & Kliegl, 2017).

A few studies employed retention intervals ranging from 24 hrs to one week between study (plus practice phase) and the final test, using selective retrieval practice without feedback as practice type. Results were mixed, with several studies reporting a neutral effect for the unpracticed material (Abel & Bäuml, 2014; MacLeod & Macrae, 2001), some other studies reporting a detrimental effect (Garcia-Bajos et al., 2009; Storm et al., 2012), and a few studies reporting a beneficial effect (Chan, 2009; Chan et al., 2006). All of these studies, which included a wide range of stimulus materials, found beneficial effects for the practiced material itself. Using a 12-hrs retention interval, Baran et al. (2010) reported a beneficial effect for the practiced items and a detrimental effect for the unpracticed items when selective practice occurred through retrieval practice plus feedback. No study has yet been reported with selective restudy as practice type and the use of a longer retention interval.

Several studies conducted both selective practice and the final test at a longer lag of 24 or 48 hrs after study, with the test following shortly upon practice. Using selective retrieval without feedback as practice type, practice again proved beneficial for recall of the practiced material (Bäuml & Wallner, 2020; MacLeod & Macrae, 2001). More importantly, beneficial effects arose also for the unpracticed material when lists of unrelated words or prose passages were employed as study material (Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014). For categorized lists, similar beneficial effects for the unpracticed material were reported when the practiced and unpracticed items belonged to different categories, whereas detrimental effects were reported when the practiced and unpracticed items shared the same categories (Bäuml & Wallner, 2020; MacLeod & Macrae, 2001). In the only study we know of that used lagged selective restudy as practice type, beneficial effects on the unpracticed material were reported with lists of unrelated items, albeit of a smaller magnitude than after selective retrieval practice (Bäuml & Dobler, 2015). No study has yet examined the effects of lagged selective retrieval practice plus feedback.

Studies on Nonselective Practice

Further information on the roles of practice type and practice timing for final recall of studied material when practice takes place during a constant, longer retention interval can be found in studies regarding retrieval practice effects. The typical finding here is that retrieval practice is more beneficial for final recall than is restudy (Karpicke & Roediger, 2008; Roediger & Karpicke, 2006). Because practice is nonselective in these studies, only the role of practice for the practiced material can be examined. Most of these studies conducted practice immediately after study and compared the effects of retrieval practice without feedback, retrieval practice plus feedback, and restudy conditions – with a few studies including no-practice control conditions as well. Study materials varied over a wide range of conditions (for details, see Karpicke, 2017). With retention intervals of one to seven days between study (plus practice phase) and final test, frequent findings were improved recall after retrieval practice with or without feedback relative to restudy conditions (Carrier & Pashler, 1992; Roediger & Karpicke, 2006), and an advantage of retrieval practice plus feedback over retrieval practice without feedback (Butler & Roediger, 2008; Kang et al., 2007). Kang et al. (2007) also reported increased retention after restudy compared to a no-practice control condition.

A few other studies on nonselective practice also varied practice timing while keeping the retention interval between study and the final test constant. Glover (1989, Experiment 2) examined the effect of retrieval practice without feedback relative to a no-practice control condition on final recall of studied prose material and placed practice either at the beginning or in the middle of a four-day retention interval. Results showed a benefit of lagged over immediate retrieval practice. Keeping the retention interval almost constant and employing face-name pairs as study material, Pyc et al. (2014, Experiment 4) used a short-lag (10 min) and a long-lag (24 hrs) practice condition and compared final recall in restudy and retrieval practice plus feedback conditions. The final test took place one week after practice and showed a higher benefit of retrieval practice over restudy after lagged than after immediate practice. For restudy as practice type, a numerical superiority of late over early practice of prose material was observed for retention intervals of both 48 hrs (Rawson, 2012) and one week (Greving & Richter, 2019).[1]

The Present Study

Returning to the fictional student's introductory question about how and when to selectively practice material to achieve the best results at test, research to date has obviously provided us with fragmentary information only. Studies have mostly focused on early or late selective retrieval practice without feedback or on various forms of early nonselective practice. Despite a few exceptions from this pattern, as a whole the studies seem insufficient at this point to provide clear-cut suggestions on which combination of practice type and practice timing within a constant longer (2-days) retention interval should prove best to remember as much of the practiced and unpracticed study material as possible. This study addresses the issue, examining the effects of different combinations of practice type and practice timing on recall of previously studied (practiced and unpracticed) material.

The results of three experiments are reported in each of which participants studied a list of unrelated words and 48 hrs later were tested on all studied items. In each experiment, practice time was either early (immediately after study), middle (24 hrs after study), or late (immediately before the test). Practice type was varied across experiments: Experiment 1 employed restudy as practice type, Experiment 2 retrieval practice without feedback, and Experiment 3 retrieval practice plus feedback. In all other aspects, the three experiments were identical. Experiment 1 also contained a no-practice baseline condition that, in all three experiments, was used to evaluate the effects of selective practice on final recall of the practiced and unpracticed items. The results of the three experiments will provide important information on how the combination of practice type and practice timing influences final recall of studied material, thus hopefully coming up with first answers on the fictional student's question. In addition, the results will reveal possible differential effects of

[1] As the authors were concerned with other objectives in these studies, no statistical analysis on the possible effects of practice timing within a fixed retention interval on recall performance was reported.

practice type and practice timing on relative final recall of the practiced and unpracticed material.

Experiment 1

Method

Ethical Considerations

All reported studies were carried out in accordance with the provisions of the World Medical Association Declaration of Helsinki.

Participants

156 students of Regensburg University participated in the experiment (M= 22.54 years, range = 18-31 years, 117 female, 39 male). They were equally distributed across the three between-participants conditions, resulting in n=52 participants in each condition. We determined the desired sample size based on reported effect sizes in prior work on the effects of selective retrieval, counterbalancing purposes, and the results of an analysis of test power conducted with the G*Power program (version 3, Faul et al., 2007). For this analysis, we set alpha at .05, power at .95, and d at 0.65. All participants were fluent in German and received monetary reward or course credit for participation.

Materials

Two lists of items were employed as study material, each consisting of 20 unrelated concrete German nouns. The lists were compiled from prior studies on the effects of selective retrieval and selective restudy (Bäuml & Samenieh, 2012; Wallner & Bäuml, 2017). Within each list, each item had a unique initial letter. For each of the two lists, two item sets were constructed. For the one set, ten items of the list were randomly selected, serving as practiced items for one half of the participants and as unpracticed items for the other half. The list's remaining ten items defined the second item set and similarly served as unpracticed items for one half of the participants and as practiced items for the other half. The distinction between item types was unknown to the participants during study.

Design

The experiment had a 3 $\times$ 2 mixed factorial design. practice timing (early practice, middle practice, late practice) was manipulated between participants, whereas item type (practiced, unpracticed) was manipulated within participants. Participants were asked to restudy half of the studied items right after study, after 24 hrs, or right before the test. Assignments to practice conditions, lists, and list sets were counterbalanced across participants.

Procedure

The whole experiment was conducted via individual zoom meetings in sound and vision for each participant during which the experimenter shared the screen. All answers were given orally by the participant and recorded by the experimenter. Participants in the early restudy condition and in the late restudy condition had two meetings via zoom, participants in the middle restudy condition had three meetings via zoom. Each first zoom meeting started with the study phase. In each of two study cycles, the 20 items of one list were exposed individually in the center of the computer screen for 5 s each and in a random order. The study phase was followed by a 2-min distractor phase, in which participants solved arithmetic problems as a recency control.

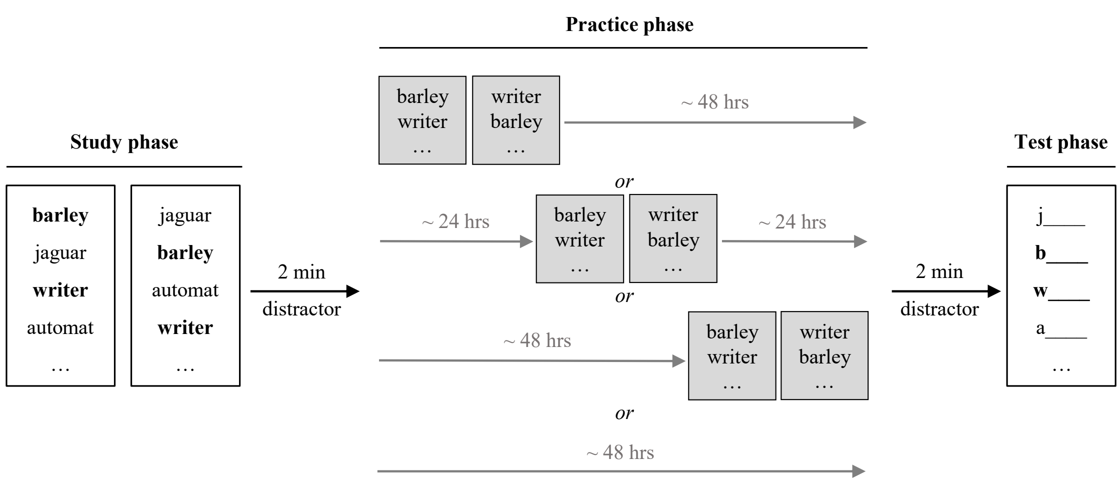

Figure 1: Procedure and conditions employed in Experiment 1. Participants studied a list of words in two successive study cycles. On the final test 48 hrs later, they were asked to recall all studied words, with the words' unique initial letters serving as retrieval cues. Restudy of half of the words was conducted in two further study cycles and took place immediately (early practice), 24 hrs (middle practice), or 48 hrs (late practice) after initial study. In an additional baseline condition, participants did not practice any items between study and test. Practiced items are depicted in bold letters.

While participants in the middle and late restudy conditions were dismissed at this point, participants in the early restudy condition were asked, prior to dismissal, to selectively restudy 10 of the previously studied items. The items were again exposed successively for 6 s each and in two successive cycles, each with its own random order. Participants were then also dismissed and asked to join the next meeting 48 hrs (+/-2 hrs) later. Participants in the middle restudy condition joined their second meeting 24 hrs (+/-1 hr) after study. They restudied half of the list's items in the same way as in the early restudy condition; the third meeting was planned another 24 hrs (+/-1 hr) later. Participants in the late restudy condition joined their second meeting 48 hrs (+/-2 hrs) after study, which started with the same two restudy cycles as in the early and middle restudy conditions. Independently of condition and prior to the test phase, verbal distractor tasks were administered for 2 min in which participants were asked to find relational similarities between word pairs. At test, all participants were presented with the initial letters of all 20 – practiced and unpracticed – items successively and in a random order for 6 s each. Because this study was equally interested in the recall of practiced and unpracticed items, we opted for this testing procedure instead of testing one item type before the other (see Figure 1 for an illustration of the experiment's procedure and conditions).

Additional Baseline Condition

Another 52 students (M = 21.58 years, range: 18-30 years, 12 male, 40 female) took part in an additional, fourth experimental condition, which, in both material and procedure, was identical to the three practice conditions with the only exception that there was no practice between study and test at all. We counterbalanced a list's same two item sets across participants as were employed in the other three experimental conditions. Mean recall then served as a baseline in all three experiments of this study to measure the extent to which possible beneficial or detrimental effects of selective practice on practiced and unpracticed items would arise.

Results

Final-test results for total recall (i.e., recall combined over practiced and unpracticed items), recall of the practiced items, and recall of the unpracticed items are depicted in Figures 2a-c, each as a function of practice timing. A 3 $\times$ 2 ANOVA with the between-participants factor of practice timing (early practice, middle practice, late practice) and the within-participant factor of item type (practiced, unpracticed) showed no main effect of practice timing, $F(2, 153) = 2.35$, $MSE = 6.72$,

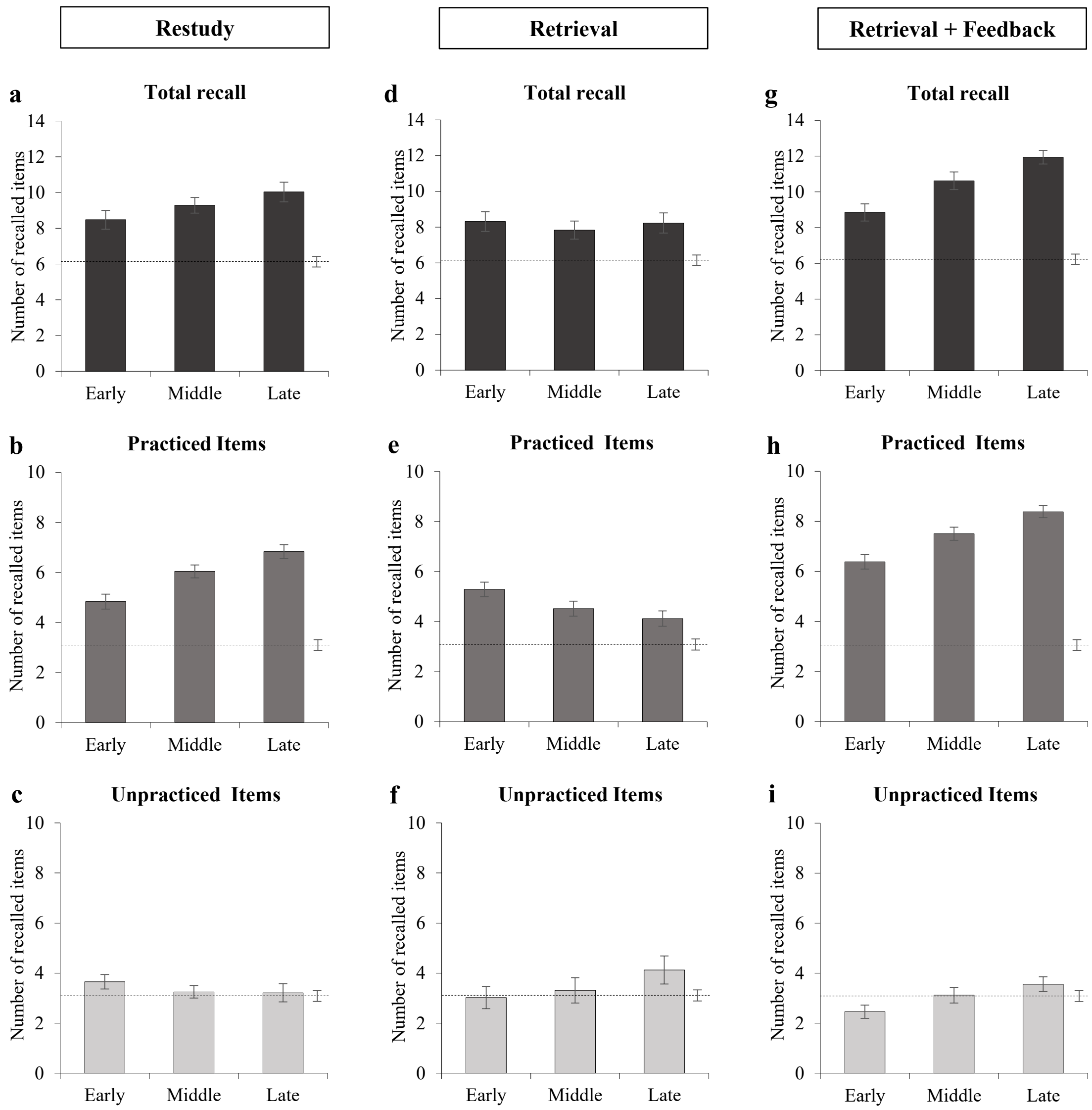

Figure 2: Results of Experiment 1 **(a-c)**, Experiment 2 **(d-f)**, and Experiment 3 **(g-i)**. For each single experiment, the figure shows final-test performance, separately as total recall, i.e., recall combined over practiced and unpracticed items, number of recalled practiced items, and number of recalled unpracticed items, each as a function of practice timing (early practice, middle practice, late practice). The dashed line represents the no-practice baseline condition. Error bars represent +/- 1 standard error.

$p = .099$, $\eta^2 = 0.03$, suggesting that practice timing did not influence total recall. In contrast, there were a significant main effect of item type, $F(1, 153) = 217.99$, $MSE = 2.28$, $p < .001$, $\eta^2 = 0.59$, with higher recall of the practiced than the unpracticed items, and a significant interaction between the two factors, $F(2, 153) = 17.58$, $MSE = 2.28$, $p < .001$, $\eta^2 = 0.19$, alluding to differential effects of practice timing on the recall of the practiced and unpracticed items. Indeed, while the unifactorial follow-up ANOVA for the practiced items revealed a significant effect, $F(2, 153) = 12.72$, $MSE = 4.15$, $p < .001$, $\eta^2 = 0.14$, with an increase in recall from early to late practice, the unifactorial follow-up ANOVA for the unpracticed items suggested no influence of practice timing on recall of these items, $F(2, 153) < 1$, $MSE = 4.85$, $p = .527$, $\eta^2 = 0.01$. For all three practice timing conditions, both total recall and recall of the practiced items differed from recall in the no-practice baseline condition (total recall: all $ts(102) > 3.40$, $ps < .001$, $ds > 0.67$; recall of the practiced items: all $ts(102) > 4.74$, $ps < .001$, $ds > 0.93$), whereas recall of the unpracticed items was indistinguishable from baseline, all $ts(102) < 1.57$, $ps > .119$, $ds < 0.31$.

Discussion

Restudy was beneficial for total recall on the final test as compared to the no-practice baseline condition. Numerically, though not statistically, this beneficial effect increased from early to late practice time, suggesting that recall might benefit if restudy was done late during the retention interval. The trend for improved recall when practice was lagged was also present for recall of the practiced items, and in this case the effect was also statistically present. In contrast, practice timing did not influence recall of the unpracticed items, which, in all cases, did not differ from recall in the no-practice baseline condition. In all three practice timing conditions, restudy therefore exclusively influenced recall of the practiced items while leaving recall of the unpracticed items unaffected. The aim of Experiment 2 was to examine whether results for selective restudy would generalize to selective retrieval practice (without feedback).

Experiment 2

Method

Participants

Another 156 students of Regensburg University took part in the experiment (M= 21.96 years, range: 18-33 years, 116 female, 40 male). Sample size followed Experiment 1. Again the participants were equally distributed across the three between-participants conditions, resulting in n=52 participants in each condition. All participants spoke German fluently and received monetary reward or course credit for participation.

Materials, Design, and Procedure

Materials, design, and procedure were identical to Experiment 1, with the only difference that type of practice changed from selective restudy to selective retrieval. Immediately after study, 24 hrs after study, or 48 hrs after study, participants were asked to selectively retrieve 10 of the previously studied items. The items' unique initial letters were provided as retrieval cues for 6 s each and in two successive practice cycles, each with its own random order.

Results

Final-test results for total recall, recall of the practiced items, and recall of the unpracticed items are depicted in Figures 2d-f. A 3 $\times$ 2 ANOVA with the between-participants factor of practice timing (early practice, middle practice, late practice) and the within-participant factor of item type (practiced, unpracticed) showed no main effect of practice timing, $F(2,153) < 1$, $MSE = 6.68$, $p = .772$, $\eta^2 < 0.01$, suggesting that practice timing did not influence total recall. However, there were a main effect of item type, $F(1,153) = 48.62$, $MSE = 2.16$, $p < .001$, $\eta^2 = 0.24$, with higher recall of the practiced than unpracticed items, and a significant interaction between the two factors, $F(2,153) = 15.52$, $MSE = 2.16$, $p < .001$, $\eta^2 = 0.17$. In fact, unifactorial follow-up ANOVAs revealed opposing effects of practice timing on recall of the two types of items. Whereas for the practiced items, recall decreased from early to late practice, $F(2,153) = 3.97$, $MSE = 4.65$, $p = .021$, $\eta^2 = 0.05$, recall increased from early to late practice for the unpracticed items, $F(2,153) = 4.01$, $MSE = 4.19$, $p = .020$, $\eta^2 = 0.05$. When compared to the no-practice baseline condition, both total recall and recall of the practiced items were higher than baseline recall in all three practice conditions (total recall: all $ts(102) > 2.52$, $ps < .013$, $ds > 0.49$; recall of the practiced items: all $ts(102) > 2.76$, $ps < .007$, $ds > 0.54$). Recall of the unpracticed items was higher than baseline recall only in the late practice condition, $t(102) = 2.68$, $p = .008$, $d = 0.53$, but did not differ from baseline recall in the other two practice conditions, both $ts(102) < 1$, $ps > .516$, $ds < 0.13$.

Number of successfully recalled practiced items during selective retrieval is depicted in the upper half of Table 1, as a function of practice cycle and practice timing. A 2 $\times$ 3 ANOVA with

Table 1. *Number of successfully recalled practiced items during selective retrieval practice as a function of practice timing (early, middle, late) and practice cycle (first, second). Results are shown for both Experiment 2 and Experiment 3. (M=mean recall; SE=standard error.)*

	1st cycle		*2nd cycle*	
	M	*SE*	*M*	*SE*
Experiment 2				
Early practice	5.00	0.27	5.67	0.30
Middle practice	3.63	0.26	4.19	0.27
Late practice	3.54	0.27	3.94	0.28
Experiment 3				
Early practice	5.17	0.23	8.96	0.20
Middle practice	3.04	0.25	8.48	0.17
Late practice	3.54	0.29	8.19	0.23

the within-participant factor of practice cycle (cycle 1, cycle 2) and the between-participants factor of practice timing (early practice, middle practice, late practice) showed a main effect of practice cycle, $F(1, 153) = 55.89$, $MSE = 0.41$, $p < .001$, $\eta^2 = 0.27$, with higher recall in the second than the first practice cycle, which suggests hypermnesia. There was also a significant main effect of practice timing, $F(2, 153) = 10.60$, $MSE = 7.53$, $p < .001$, $\eta^2 = 0.12$, with lower recall at later than earlier practice timings, which indicates typical time-dependent forgetting over the course of the 48 hrs retention interval. There was no interaction between the two factors, $F(2, 153) = 1.15$, $MSE = 0.41$, $p = .321$, $\eta^2 = 0.02$.

Discussion

Retrieval practice was beneficial for total recall on the final test as compared to the no-practice baseline condition. Critically, this beneficial effect was unaffected by practice timing with underlying opposing effects of practice timing on the practiced and unpracticed items. Indeed, whereas recall of the practiced items decreased from early to late practice, recall of the unpracticed items increased across the three practice timing conditions. As another difference between the two types of items, the results showed that recall of the practiced items was higher than baseline recall in all three practice timing conditions, while recall of the unpracticed items was above baseline level in the late practice timing condition only. When compared to the results of Experiment 1, the results of Experiment 2 thus suggest different roles of practice timing and item type for final recall after selective retrieval practice than after selective restudy. Experiment 3 was aimed to examine how selective retrieval practice plus feedback would influence recall in the same three practice timing conditions.

Experiment 3

Method

Participants

Another 156 students of Regensburg University took part in the experiment (M= 22.51 years, range: 18-30 years, 131 female, 25 male). Sample size followed Experiments 1 and 2. Again the participants were equally distributed across the three between-participants conditions, resulting in n=52 participants in each condition. All participants spoke German fluently and received monetary reward or course credit for participation.

Materials, Design, and Procedure

Materials, design, and procedure were identical to Experiment 2, with the only difference that retrieval practice comprised feedback. That is, during retrieval practice, participants not only tried to selectively retrieve the 10 practiced items but, after the attempted recall of each single item, were provided with the correct answer on the computer screen for 1.5 s, irrespective of their response. In all other aspects, Experiment 3 was identical to Experiment 2.

Results

Final-test results for total recall, recall of the practiced items, and recall of the unpracticed items are depicted in Figures 2g-i. A 3 $\times$ 2 ANOVA with the between-participants factor of practice timing (early practice, middle practice, late practice) and the within-participant factor of item type (practiced, unpracticed) showed a main effect of practice timing, $F(21, 153) = 11.68$, $MSE = 5.37$, $p < .001$, $\eta^2 = 0.13$, with total recall increasing from early to late practice timing, as well as a main effect of item type, $F(1, 153) = 573.73$, $MSE = 2.61$, $p < .001$, $\eta^2 = 0.79$, indicating higher recall of practiced than unpracticed items. The interaction between the two factors was not significant, $F(2, 153) = 2.04$, $MSE = 2.61$, $p = .134$, $\eta^2 = 0.03$. Both total recall and recall of the practiced items exceeded baseline recall in all three practice timing conditions (total recall: all $ts(102) > 4.18$, $ps < .001$, $ds > 0.82$; recall of practiced items: all $ts(102) > 9.31$, $ps < .001$, $ds > 1.83$), whereas recall of the unpracticed items did not differ from baseline recall in any of the practice timing conditions, all $ts(102) < 1.85$, $ps > 0.67$, $ds < 0.36$.

Number of successfully recalled practiced items during selective retrieval is depicted in the lower half of Table 1, as a function of practice cycle and practice timing. A 2×3 ANOVA with the within-participant factor of practice cycle (cycle 1, cycle 2) and the between-participants factor of practice timing (early practice, middle practice, late practice) showed a main effect of practice cycle, $F(1, 153) = 1044.95$, $MSE = 1.60$, $p < .001$, $\eta^2 = 0.87$, with much higher recall in the second than the first practice cycle, suggesting beneficial effects of feedback. There was also a significant main effect of practice timing, $F(2, 153) = 13.84$, $MSE = 3.97$, $p < .001$, $\eta^2 = 0.15$, with lower recall at later than earlier practice timings, indicating typical time-dependent forgetting. Additionally, the interaction between the two factors was significant, $F(2, 153) = 11.13$, $MSE = 1.60$, $p < .001$, $\eta^2 = 0.13$, with more time-dependent forgetting in the first than the second practice cycle, which points to hypermnesia and a reduction of time-dependent forgetting through feedback.

Additional Between-Experiments Analyses

The above analyses suggest qualitative differences between the effects of practice type on final-test performance as they were studied in Experiment 1 (restudy), Experiment 2 (retrieval practice without feedback), and Experiment 3 (retrieval practice plus feedback). First, while total recall increased from early to late practice when retrieval practice plus feedback was employed – and did so at least numerically also after restudy –, total recall was unaffected by practice timing after retrieval practice without feedback. Second, recall of practiced items increased across practice times when restudy or retrieval practice plus feedback were employed, but decreased after retrieval practice without feedback. Third, recall of unpracticed items increased from early to late practice when retrieval practice with or without feedback were employed, but was unaffected when restudy was conducted. We examined whether these qualitative differences between practice types were also statistically present by conducting a 3 $\times$ 3 $\times$ 2 ANOVA with the between-participants factors of experiment (restudy, retrieval practice without feedback, retrieval practice plus feedback) and practice timing (early practice, middle practice, late practice) and the within-participant factor of item type (practiced, unpracticed).

The ANOVA revealed main effects of experiment, $F(2, 459) = 17.16$, $MSE = 6.26$, $p < .001$, $\eta^2 = 0.07$, with highest recall after retrieval practice plus feedback and lowest after retrieval practice without feedback, practice timing, $F(2, 459) = 7.27$, $MSE = 6.26$, $p = .001$, $\eta^2 = 0.03$, with highest recall after late and lowest after early practice, and item type, $F(1, 459) = 719.66$, $MSE = 2.35$, $p < .001$, $\eta^2 = 0.61$. In particular, the analysis showed significant interactions between experiment and practice timing, $F(4, 459) = 2.78$, $MSE = 6.26$, $p = .027$, $\eta^2 = 0.02$, and experiment and item type, $F(2, 459) = 86.6$, $MSE = 2.35$, $p < .001$, $\eta^2 = 0.27$, but no significant interaction between practice timing and item type, $F(2, 459) = 1.35$, $MSE = 2.35$, $p = .259$, $\eta^2 < 0.01$. There was also a significant three-way interaction, $F(4, 459) = 16.13$, $MSE = 2.35$, $p < .001$, $\eta^2 = 0.12$. The interaction between experiment and practice timing together with the three-way interaction mirror the qualitative differences in the effects of practice types emphasized in the previous paragraph.

A major goal of the present study was to identify which combination of practice type and practice timing would yield the highest final recall performance. The results of Experiments 1-3 indicate that, for both retrieval practice plus feedback and restudy, recall was best after late practice, whereas for retrieval practice without feedback recall was unaffected by practice timing (see also Figure 2). To identify the "best" practice combination, we therefore focused on the late practice conditions and conducted a 3 $\times$ 2 ANOVA with the factors of experiment and item type to analyze which practice type induced the highest recall performance.

Above all, the ANOVA showed a main effect of experiment, $F(2, 153) = 13.53$, $MSE = 6.62$, $p < .001$, $\eta^2 = 0.15$, with total recall being higher after (late) retrieval practice plus feedback than after (late) restudy, $t(102) = 2.85$, $p = .005$, $d = 0.56$, and higher after (late) restudy than after (late) retrieval practice without feedback, $t(102) = 2.30$, $p = .024$, $d = 0.45$. There were also a main effect of item type, $F(1, 153) = 207.77$, $MSE = 2.97$, $p < .001$, $\eta^2 = 0.58$, and a significant interaction, $F(2, 153) = 55.15$, $MSE = 2.97$, $p < .001$, $\eta^2 = 0.42$. Like total recall, recall of practiced items was higher after (late) retrieval practice plus feedback than after (late) restudy, $t(102) = 4.19$, $p < .001$, $d = 0.82$, and higher after (late) restudy than after (late) retrieval practice without feedback, $t(102) = 6.46$, $p < .001$, $d = 1.27$. For unpracticed items, no significant differences emerged between practice types, $F(2, 153) = 1.95$, $MSE = 5.55, p = .146$, $\eta^2 = 0.03$. Late retrieval practice plus feedback therefore reflected the "best" practice combination in the present set of experiments.

General Discussion

We set out to answer a fictional student's question of how and when to selectively practice previously studied subject matter to maximize retention at a test two days after study. With regard to total recall, two main tendencies emerged from the present experiments. First, mean recall was best after retrieval practice plus feedback followed by restudy and retrieval practice without feedback; and, second, mean recall increased from early to middle and late practice. Above all, the results clearly demonstrated that total recall was best after late retrieval plus feedback. In none of the single practice timing conditions did retrieval practice without feedback induce higher total recall than the other two practice types and, for middle and late practice, even created lower recall than either restudy or retrieval practice plus feedback.[2] The results therefore indicate that, in order to maximize total recall, a student should prefer retrieval practice plus feedback and a rather late practice for practicing studied information.

Analysis of recall of the practiced and unpracticed items revealed that the results for total recall were largely driven by the recall of the practiced items, which showed the same two main tendencies as emerged for total recall. Consistently, mean recall of the practiced items was highest for

[2]Practice type did not influence total recall after early practice, $F(2, 153) < 1, MSE = 12.4, p = .729, \eta^2 < 0.01$, but did so after middle and late practice, both $Fs(2, 153) > 8.5$, $MSEs > 11.9$, $ps < .001$, $\eta^2 > 0.10$. Indeed, after both middle and late practice, total recall was lower after retrieval without feedback than in the other two practice type conditions, all $ts(102) > 2.19$, $ps < .031$, $ds > 0.43$.

retrieval practice plus feedback and lowest for retrieval practice without feedback as well, and improved from early to late practice. In particular, recall of the practiced items was also best after late retrieval practice plus feedback. In contrast, recall of the unpracticed items was relatively low and did not much vary with practice type. Although, in the two retrieval practice conditions, recall of the unpracticed items did also improve from early to late practice, only for late retrieval practice without feedback did recall of these items differ from baseline level and show a beneficial effect. Preferring late retrieval practice without feedback as practice format would come with costs, however, because recall of practiced items under this condition was rather low relative to the restudy and retrieval practice plus feedback conditions.[3]

The low recall level for the practiced items in the late retrieval practice without feedback condition likely was caused by the low success rates for practiced items in this condition (see Table 1), as low success rates during practice typically induce low recall levels at test (see Kang et al., 2007). In contrast, in both the late restudy and the late retrieval-practice plus feedback conditions, *all* practiced items should have gained additional strength in memory representation - be it through rereading or feedback -, thus ending up at higher mean final recall levels than in the late retrieval practice without feedback condition. This view on the observed differences in practiced items' recall levels mirrors basic assumptions on retrieval practice effects as they are, for instance, incorporated in the bifurcation model (Kornell et al., 2011) and its extension to time-lagged retrieval practice plus feedback (Pastötter & Bäuml, 2016).

Relation to Prior Work on Selective and Nonselective Practice

Several of the present results fit with prior work on the effects of selective practice. For instance, we replicated prior work by finding beneficial effects of selective retrieval practice without feedback for final recall of the practiced material and neutral effects for recall of the unpracticed material when practice occurred shortly after study and the retention interval was long (Abel & Bäuml, 2014; MacLeod & Macrae, 2001; but see Chan et al., 2006, or Storm et al., 2012); similarly, we found beneficial effects for the unpracticed items when retrieval practice without feedback was lagged

[3] Practiced and unpracticed items showed exactly the same recall level in this condition ($M = 4.12$ items).

and occurred immediately before the test (Bäuml & Dobler, 2015; Bäuml & Schlichting, 2014); and we found a higher benefit for unpracticed items after late retrieval practice without feedback than after late restudy (Bäuml & Dobler, 2015). In contrast to Baran et al. (2010), we did not find a detrimental effect for unpracticed items after early retrieval practice plus feedback. The reason for this discrepancy may be the difference in employed retention intervals (12 hours in Baran et al. versus 2 days in the present study), because the detrimental effect for the unpracticed items has often been found to decrease with the length of the retention interval (see Bäuml & Kliegl, 2017).

Several of the present results also fit with prior work on the effects of nonselective practice when the influence of different practice timings on later recall of studied material was examined. Pyc et al. (2014, Experiment 4) compared the effects of restudy and retrieval practice plus feedback after a 10-min and a 24-hrs lag, and, seven days after practice, found a higher benefit of retrieval practice over restudy after the long lag, which is consistent with the present results. Using nonselective restudy as practice type, both Rawson (2012) and Greving and Richter (2019) found a numerical superiority of late over early practice for retention intervals of 48 hrs and one week, which also fits with the present results. In contrast, the results of at least one study disagree with the present findings. Glover (1989, Experiment 2) found a benefit of nonselective retrieval practice without feedback over a no-practice control condition when practice was in the middle but not when it was at the beginning of a 4-days retention interval. However, these results also deviate from those of many other studies in the literature, which, like present Experiment 2, reported a beneficial effect of retrieval practice over a no-practice baseline when retrieval practice without feedback was early (Kang et al., 2007; Spitzer, 1939).

Possible Limitations of the Present Study

In this study, we used short lists of unrelated words as study material to address the question of how and when to selectively practice previously studied information to maximize long-term retention. This choice of study material raises the question of whether the present results will extend to more complex study material with its typically larger amount of information, and arise also outside the lab, like in educational settings, when scholars or students may practice material in order to maximize retention for a later exam. At least to date, retrieval practice effects have often been found to generalize from relatively short word lists to more applied settings (e.g. Carpenter et al., 2009; McDaniel et al., 2007). Another issue of generalizability is whether the results of the present experiments will still hold when the retention interval between study and test is prolonged further and other, possibly more fine-graded intermediate practice schedules are employed. Moreover, students may be concerned about longer-term retention, for instance, retention of the studied material several weeks after the test. Indeed, knowing whether the ideal time of practice is “always” shortly before the test regardless of length of retention interval and regardless of whether the interest is more on shorter- or on longer-term retention is of high relevance for many applied situations. Future studies may examine these important issues.

Conclusions

This study investigated how and when during a longer retention interval selective practice of studied material should take place to maximize later test performance. The results of three experiments show that, on average, recall is best after retrieval practice plus feedback and worst after retrieval practice without feedback, and that recall improves from early to late practice. Above all, the findings indicate that final recall can be maximized through late retrieval practice plus feedback and that this effect is largely driven by high recall levels for the practiced items. This holds while the results also point to pronounced differences between practice types regarding the effects of practice timing and the relative contributions of practiced and unpracticed material to recall performance.

Author Contributions

Lisa Wallner: Data curation, Formal analysis (lead), Investigation, Methodology, Visualization (lead), Writing – original draft. **Anna T. Nickl:** Formal analysis (supporting), Visualization (supporting), Writing – original draft, review & editing. **Karl-Heinz T. Bäuml:** Conceptualization, Methodology, Resources, Supervision, Writing – review & editing.

References

Abel, M., & Bäuml, K.-H. T. (2014). The roles of delay and retroactive interference in retrieval-induced forgetting. *Memory & Cognition, 42*(1), 141–150.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1063-1087.

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic Bulletin & Review, 7*(3), 522–530.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102*(1), 68–100.

Baran, B., Wilson, J., & Spencer, R. M. C. (2010). REM-dependent repair of competitive memory suppression. *Experimental Brain Research, 203*(2), 471–477.

Bäuml, K.-H. T., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory & Cognition, 32*(4), 610–617.

Bäuml, K.-H. T., & Dobler, I. M. (2015). The two faces of selective memory retrieval: Recall specificity of the detrimental but not the beneficial effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(1), 246–253.

Bäuml, K.-H. T., & Kliegl, O. (2017). Retrieval-induced remembering and forgetting. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (2., pp. 27–51). Oxford: Academic Press.

Bäuml, K.-H. T., & Samenieh, A. (2012). Influences of part-list cuing on different forms of episodic forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(2), 366-375.

Bäuml, K.-H. T., & Schlichting, A. (2014). Memory retrieval as a self-propagating process. *Cognition, 132*(1), 16–21.

Bäuml, K.-H. T., & Wallner, L. (2020). Selective retrieval in categorized lists: detrimental, neutral, and beneficial e□ects on nonretrieved items. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(7), 1372–1386.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*(3), 604–616.

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

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition, 20*(6), 633–642.

Chan, J. C. K. (2009). When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. *Journal of Memory and Language, 61*(2), 153–170.

Chan, J. C. K., McDermott, K. B., & Roediger, H. L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 135*(4), 553–571.

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1403–1414.

Erdman, M. R., & Chan, J. C. K. (2013). Providing corrective feedback during retrieval practice does not increase retrieval- induced forgetting. *Journal of Cognitive Psychology, 25*(6), 692–703.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A exible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research*

Methods, 39(2), 175-191.

Garcia-Bajos, E., Migueles, M., & Anderson, M. C. (2009). Script knowledge modulates retrieval-induced forgetting for eyewitness events. *Memory, 17*(1), 92–103.

Glover, J. A. (1989). The "testing" phenomenon: not gone but nearly forgotten. *Journal of Educational Psychology, 81*(3), 392–399.

Greving, C. E., & Richter, T. (2019). Distributed learning in the classroom: Effects of rereading schedules depend on time of test. *Frontiers in Psychology, 9*, Article 2517.

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*(4–5), 528–558.

Karpicke, J. D. (2017). Retrieval-based learning: A decade of progress. In J. H. Byrne (Ed.) *Learning and memory: A comprehensive reference* (2., pp. 487–514). Oxford: Academic Press.

Karpicke, J. D., & Roediger, H. L. (2008). The Critical Importance of retrieval for learning. *Science, 319*(5865), 966–968.

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*(2), 85–97.

MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science, 12*(2), 148–152.

McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review, 14*(2), 200–206.

Pastötter, B. & Bäuml, K.-H. T. (2016). Reversing the testing effect by feedback: Behavioral and electrophysiological evidence. *Cognitive, Affective, and Behavioral Neuroscience, 16*(3), 473-488.

Pyc, M. A., Balota, D. A., McDermott, K. B., Tully, T., & Roediger, H. L. (2014). Between-list lag effects in recall depend on retention interval. *Memory & Cognition, 42*(6), 965–977.

Rawson, K. A. (2012). Why do rereading lag effects depend on test delay? *Journal of Memory and Language, 66*(4), 870–884.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255.

Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology, 30*(9), 641–656.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2012). On the durability of retrieval-induced forgetting. *Journal of Cognitive Psychology, 24*(5), 617–629.

Wallner, L., & Bäuml, K.-H. T. (2017). Beneficial effects of selective item repetition on the recall of other items. *Journal of Memory and Language, 95*, 159–172.

Author Note

Research Disclosure Statements

All dependent variables or measures that were analyzed for this article's target research question have been reported in the Methods sections. Also all levels of all independent variables or all predictors and manipulations, whether successful or failed, have been reported in the Method sections. No observations were excluded.

Open Practices Statement

Neither of the experiments reported in this article was formally preregistered. However, the study materials employed in the present experiments as well as the data from the single experiments are available on the Open Science Framework ($https://osf.io/53s2w/$). Further requests for the data or materials can be sent via email to the corresponding author at [karl-heinz.baeuml@ur.de].

Software

All experiments reported in this manuscript were implemented using the videotelephony software program Zoom (Zoom Video Communications) and the software PowerPoint 1909 (Microsoft Corporation). The softwares were run on standard desktop computers by the experimenters with the operating system Windows 10 (Microsoft, Redmond, WA). The equipment of the participants was not controlled. All data were analyzed using IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp., Armonk, NY) and G*Power 3.1 (Faul et al., 2007).

Declaration of Conflicting Interests

The authors have no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.