Why do learners ignore expected feedback in making metacognitive decisions about retrieval practice?

Thomas C. Toppino [1] · **Kelsey A. Heslin** [1] · **Taylor M. Curley** [1] · **Michael K. Jackiewicz** [1] · **Colin S. Flowers** [1] · **Heather-Anne Phelan** [1]

Abstract

We report two experiments investigating why learners, in making metacognitive judgments, often seem to ignore or otherwise fail to appreciate that feedback following retrieval practice provides a restudy opportunity. Learners practiced word pairs for a final cued-recall test by studying each pair initially, making a judgment of learning (JOL), and then deciding whether to practice the pair again after a short or long spacing interval, or not at all. For different groups in Experiment 1, additional practice involved restudying, retrieval practice without feedback, or retrieval practice with feedback (the full pair). We used procedures (long feedback duration and covert retrieval practice) designed to rule out the possibility that feedback is ignored because it is usually brief or because participants' choices are influenced by a desire to look good by performing well on overt practice tests. In the relearning condition, learners preferred a long spacing interval for items at all JOL levels. Despite the feedback duration and the covert retrieval practice, learners in *both* retrieval-practice conditions preferred a short spacing interval for hard, low-JOL items and a long spacing interval for easy, high-JOL items, even though this may not be an effective strategy when feedback is provided. In Experiment 2, instructions framed feedback either as a presentation of the correct answer or as a restudy opportunity preceded by retrieval practice. Framing feedback as a restudy opportunity markedly changed the choices learners made. Apparently, the restudy function of feedback does not occur to learners unless they are specifically alerted to it.

Keywords Metamemory · Metacognition · Retrieval practice · Spacing| · Feedback

Introduction

Metacognition is thought to play a central role in self-regulated learning. There is agreement among researchers that learners use their assessments of their current state of knowledge and their personal theories of how learning and memory processes work to control their study strategies in potentially adaptive ways (e.g., Bjork et al., 2013; Nelson & Narens, 1990). The present paper is concerned with the effect of anticipated feedback when self-regulated learning involves retrieval practice.

It is well established that retrieval practice leads to greater learning than repeated studying (e.g., Karpicke & Roediger, 2008; Roediger & Butler, 2011; Rowland, 2014). According to Roediger and Karpicke (2006a), the act of retrieval per se may directly produce a substantial increase in the strength of memory or it may influence memory through a more indirect or mediated mechanism. For example, there is evidence that retrieval practice results in enhanced learning from a subsequent study opportunity compared to studying without prior retrieval practice, a phenomenon known as test-potentiated learning (e.g., Arnold & McDermott, 2013; Izawa, 1966; Soderstrom & Bjork, 2014). In spite of the powerful effect of retrieval practice, however, learners often fail to appreciate that it yields a greater effect on learning than studying by itself, and they even erroneously ascribe greater benefits to studying alone (Karpicke & Roediger, 2008; Roediger & Karpicke, 2006b).

Several studies have surveyed learners about the strategies they use in real-life learning situations (e.g., Hartwig & Dunlosky, 2012; Karpicke et al., 2009; Kornell & Bjork, 2007; Yan et al., 2014). Learners consistently reported restudying with high frequency, whereas the frequency with which they reported engaging in retrieval practice is quite

✉ Thomas C. Toppino
thomas.toppino@villanova.edu

1   Department of Psychological and Brain Sciences, Villanova
    University, 800 Lancaster Avenue, Villanova, PA 19085, USA

variable. A more fine-grained assessment (Kuhbander & Emmerdinger, 2019) indicated that students often report mixed strategies, targeted for subunits of information. They also are more prone to restudy early and to test themselves later in the learning process. Across surveys, relatively few students who included retrieval practice as a strategy reported doing so because they perceived it to have a superior effect on learning. Rather, they were most likely to choose retrieval practice to diagnose their level of learning. Interestingly, a survey by Morehead et al. (2016) found that a solid majority of college instructors also endorsed the diagnostic function of retrieval practice (68%) over its effect on learning (19%).

Early laboratory studies of self-regulated learning in which participants could choose whether to restudy or engage in retrieval practice (i.e., take a practice test) seemed generally consistent with the survey results. Although learners did choose retrieval practice, they typically under-utilized the strategy (e.g., Karpicke, 2009; Kornell & Son, 2009) or used it inefficiently by reserving it until relatively late in the learning process, after significant learning already had taken place (Karpicke, 2009). However, Toppino et al. (2018) pointed out that a reluctance to use retrieval practice when no feedback was provided may have been appropriate in the previously mentioned studies because the interval between initial studying and retrieval practice was relatively long. Without feedback, learners must be successful on the practice test to benefit from retrieval practice (e.g., Kang et al., 2007; Roediger & Butler, 2011; Storm et al., 2014). In contrast, when the practice test is followed by feedback that re-presents the information being learned (e.g., in learning word pairs, re-presentation of the full pair following a cued-recall test), retrieval practice might be expected to be chosen more often, even when the conditions are unfavorable for retrieval. In this case, successful retrieval is not necessary for a learner to gain from taking a practice test. Feedback provides a restudy opportunity.

Pashler and his colleagues (Pashler et al., 2003; Pashler et al., 2005) assessed the role of feedback as a learning opportunity in several experiments in which participants studied to-be-learned pairs, took two practice tests and then a final test. When items were recalled correctly on the practice tests, Pashler et al. (2005) found that the presence or absence of immediate corrective feedback did not affect final test performance even after a long retention interval (see also Toppino & Pagano, 2020). However, other research has found that delayed feedback can be beneficial following correct responses (e.g., Smith & Kimball, 2010). When participants made practice-test errors, Pashler et al. (2005) found that presenting the correct answer (intact pair) as immediate feedback produced far better final test performance than simple right/wrong feedback or no feedback at all. In addition, when immediate corrective feedback was presented and the spacing between practice tests was varied, Pashler et al. (2003) found

that longer inter-test lags produced better final recall and that the spacing function was roughly parallel for items that had and had not been recalled correctly on the practice tests. Finally, as noted previously, taking a test prior to restudying potentiates learning, enhancing it beyond the level that would be expected from studying alone (e.g., Arnold & McDermott, 2013; Izawa, 1966; Soderstrom & Bjork, 2014). These findings clearly indicate that corrective feedback functions as a restudy opportunity. Unfortunately, it is not clear that learners always appreciate the learning opportunity afforded by practice-test feedback.

One indication of this lack of appreciation comes from a study by Kornell and Rhodes (2013). They set out to assess mechanisms proposed to underlie the effect of delaying judgments of learning, although that is not our interest in the results. Participants received an initial study trial, followed by either a restudy trial, an interim test without feedback, or an interim test with feedback. Then, they received a prompt to make a JOL, indicating the likelihood of recalling the item later. Lastly, participants took a final cued-recall test. The results of primary interest in the present context involve the two test groups. Final recall was much higher when the test was followed by feedback, an expected finding because feedback constituted an extra study opportunity that was not enjoyed when feedback was not presented. However, the two test groups did not differ on JOLs, even though the JOL for the feedback group was made after participants had received an additional study opportunity in the form of feedback. It appears that learners based their JOLs on the results of the prior interim test and largely discounted the learning that occurred from feedback. In a final experiment, the JOL decision was replaced with a decision either to study the item again before the final test or to drop it from further study, although no additional study opportunity actually was provided before the final test. Again, recall was better when the interim test was followed by feedback compared to when it was not, attesting to the significant learning afforded by feedback. However, learners in the feedback condition did not choose to drop more items than did those in the no-feedback condition. This may indicate that learners receiving feedback underestimated their likelihood of getting items correct on the final test, perhaps because they did not take into account the learning that was enabled by feedback.

Using very similar methodology, Sitzman et al. (2016) found that, in contrast to Kornell and Rhodes' (2013) findings, JOLs after a practice test were higher when the test was followed by feedback than when it was not. Although the source of the discrepancy in results remains unclear, Sitzman et al.'s finding suggests that participants are sometimes aware of having learned from practice-test feedback. However, their JOLs were poorly calibrated, greatly underestimating final

recall. Thus, even when participants seem to realize that they learned from feedback, they may greatly discount the extent of that learning.

Further indications that learners do not always appreciate the learning potential of feedback have come from several studies of self-regulated learning in which learners chose whether to restudy or take a practice test (Toppino et al., 2018; Tullis et al., 2018) or chose whether the spacing interval between initial studying and a subsequent practice test would be short or long (Toppino & Pagano, 2020). Learners in these studies have shown little evidence that they anticipate learning from feedback when they expect it to be presented after practice tests.

In Toppino et al.'s (2018) Experiment 1, participants learned three lists of word pairs for a final cued-recall test that was not administered until practice on all three lists had been completed. When a pair was presented initially for study, participants made a JOL and decided whether to restudy the item, take a practice test, or forego further practice (the Done option). Results indicated that learners preferred to restudy when pairs were perceived to be difficult (low JOLs) and the spacing interval between item presentations was long. In contrast, they preferred retrieval practice when the pairs were judged to be easy (high JOLs) and the spacing interval was short. That is, learners preferred restudying when the likelihood of practice-test success was low and preferred retrieval practice when the chance of practice-test success was high. This appears to be an appropriate strategy when practice tests are not followed by feedback and suggests that learners may appreciate the fact that testing without feedback is only helpful if retrieval practice is successful. However, when feedback is provided so that a restudy opportunity follows a practice test regardless of retrieval success, one might expect a shift in learners' strategies such that retrieval practice might be chosen more often for harder items and longer spacing intervals. Providing feedback, however, had little effect. The anticipation of receiving feedback following a practice test had no effect at all on learners' choices on List 1. And, although feedback did increase the overall preference for retrieval practice on lists 2 and 3, anticipating feedback after a practice test did not alter the relative preference for restudying or retrieval practice as a function of JOL and spacing interval. There was no evidence of a shift in preference such that practice tests with feedback were chosen relatively more often for harder items and longer spacing intervals.

Tullis et al. (2018) also presented word pairs to be learned for a later cued-recall test. After studying an item on the initial trial, participants chose whether they would restudy or take a practice test on the item's second occurrence. They found that learners preferred retrieval practice for easy items and restudying for hard items, consistent with Toppino et al.'s (2018) results. The fact that this pattern of choices was obtained when practice tests were not followed by feedback is consistent again with the hypothesis that learners appreciate that retrieval practice must be successful for it to benefit learning. However, the pattern of choices was not altered when learners expected feedback after retrieval practice. Apparently, learners are not aware that taking a test prior to restudying potentiates the effect of restudying, but, beyond that, it remains unclear why a practice test plus a restudy opportunity is not chosen more often than a practice test alone (i.e., one without feedback).

In another recent study, Toppino and Pagano (2020) investigated learners' metacognitive control over the temporal distribution of practice tests. Participants studied word pairs for a final cued-recall test. After studying a pair initially, they made a JOL and chose whether to practice again after a short or a long spacing interval or to be done with the item, passing up the chance for further practice. Depending on the group to which participants were assigned, further practice or repetition involved restudying pairs, taking a practice test without feedback, or taking a practice test with feedback. When repetition involved restudying, learners preferred a longer spacing interval regardless of the perceived difficulty (JOL) of the item. This finding replicated previous experiments (e.g., Benjamin & Bird, 2006; Pyc & Dunlosky, 2010; Toppino et al., 2009; Toppino & Cohen, 2010) that have investigated the distribution of restudy opportunities in self-regulated learning under similar conditions (cf., Son, 2004). The preference for longer spacing when practice involves restudying suggests that learners may have some appreciation of the relative advantage of longer over shorter spacing intervals. In contrast, Toppino and Pagano found that, when repetition involved a practice test, regardless of whether it was followed by feedback, learners preferred a short spacing interval for hard, low-JOL pairs and a long spacing interval for easier, high-JOL pairs. This pattern of choices seems appropriate when practice tests do not involve feedback. Choosing a short spacing interval for hard items maximizes the likelihood of successful retrieval practice for these items, whereas choosing a long spacing interval for easy, high-JOL items provides a chance to gain the advantage of spaced practice when successful retrieval appears likely. What was most remarkable, however, was that expecting feedback after retrieval practice did not alter learners' spacing choices relative to the condition in which no feedback was expected. One might have predicted that having a restudy opportunity (i.e., feedback) in addition to a practice test would have shifted the pattern of spacing choices at least somewhat in the direction that was obtained when only restudy opportunities were available. However, even though feedback produced substantial learning as it has in many previous studies (e.g., Pashler et al., 2003, 2005), knowing that practice tests would be followed by feedback had no effect on learners' spacing choices. This constitutes a major metacognitive error.

In the present experiments, we explored the source of this metacognitive error. We adopted the paradigm used by Toppino and Pagano (2020, Experiment 1). However, whereas they were interested primarily in learners' metacognitive control over when to schedule practice as a function of JOLs and type of practice, our primary focus is on learners' metacognitive awareness of the benefits that feedback after retrieval practice has on learning.

Experiment 1

In this experiment, we explored two hypotheses about why learners seem to ignore the learning potential of anticipated feedback when making strategic decisions about using retrieval practice in self-directed learning of word pairs. First, in making strategic choices, learners may be aware of the learning benefits of feedback, but this knowledge may be overridden by implicit demands to perform well on practice tests. Learners in most studies are instructed that their choices should be guided by the goal of performing well on the final recall test, but practice tests may be more salient during learning than the relatively distant final test. Furthermore, learners are aware that their performance on practice tests is being observed and may wish to look good by performing well on those interim tests. To the extent that the desire to perform well on practice tests drives learners' strategic decisions, learners would be expected to opt for practice tests in conditions that would make successful retrieval practice likely, which is an apt description of what learners have been observed to actually do (e.g., Toppino et al., 2018; Toppino & Pagano, 2020). Under these circumstances, feedback might be expected to have little effect on learners' decisions because post-test events, such as the presence or absence of feedback, do not affect the likelihood of successful retrieval practice.

Second, learners may underestimate the learning potential of feedback due to its duration or anticipated duration. The presentation duration of the full pair is often shorter when it is presented as feedback than when it is presented on an initial study trial or on a restudy trial (e.g., Toppino et al., 2018; Toppino & Pagano, 2020). The relatively brief duration of feedback may lead participants to discount its potential usefulness as a learning opportunity.[1]

The two factors we have described (i.e., implicit demands to focus on practice tests and feedback duration) may work together, each accounting for part of learners' apparent tendency to ignore feedback. In Experiment 1, we investigated participants' strategic choices under conditions designed to

reduce or eliminate the influence of both factors. As in Toppino and Pagano's (2020) first experiment, participants studied each word pair, made a JOL, and then decided whether further practice would involve a short or a long spacing interval. They also could choose to dispense with further practice. Participants were assigned to one of three groups differing in whether additional practice entailed restudying the pair (Restudy condition), taking a practice test without feedback (Test-No Feedback condition), or taking a practice test followed by feedback (Test-Feedback condition).

To reduce implicit demands to make practice-test performance a priority, we instructed participants to respond covertly, rather than overtly, during practice tests. There is evidence that covert retrieval facilitates memory performance as much as overt retrieval for simple paired associate learning (Smith et al., 2013), although overt retrieval may have some advantage in more complex learning situations (e.g., Tauber et al., 2018). Our interest in covert retrieval was that making retrieval practice internal and unobservable should reduce or eliminate implicit pressure on participants to look good by performing well on practice tests. The participants themselves would be the only ones to know whether or not their retrieval was successful. The use of covert retrieval practice might free learners, at least somewhat, from demands to perform well on practice tests so that they would be more likely take into account other factors such as feedback.

To reduce the likelihood that participants in the present experiment would discount the learning potential of feedback due to its duration, we used a relatively long 5-s duration of feedback and equated it with the duration of other learning events in the experiment. Thus, a 5-s duration was used for the presentation of the full word pair on initial study trials, the presentation of the cue word alone during the practice test, the presentation of the full word pair as a restudy trial, and the presentation of the full word-pair as feedback.

We expected to replicate Toppino and Pagano's (2020) findings for the Restudy and the Test-No Feedback conditions. That is, we expected participants to prefer a longer spacing interval over a shorter one in the Restudy condition, regardless of an item's JOL level. In contrast, we expected participants in the Test-No Feedback condition to prefer a short spacing interval for the hard, low-JOL items and a long spacing interval for easy, high-JOL items. However, whereas Toppino and Pagano obtained the same pattern of choices for the two Test conditions, we expected a difference to emerge to the extent that covert retrieval practice and long feedback durations operated to reduce factors that may have lowered participants' attention to feedback in the past. Specifically, compared to the Test-No Feedback condition, participants in the Test-Feedback condition would be expected to choose a longer spacing interval more often, especially for low-JOL items.

---

[1] Tullis et al. (2018) allowed participants to control the duration of feedback, and it is unclear whether or how feedback duration may have been related to participants' decisions about retrieval practice.

Method

Participants

Sample size was determined by an a priori power analysis based on one of several large effects obtained by Toppino and Pagano (2020, Experiment 1), namely, a 2 × 3 within/between interaction for which η^2_p = .265. Results indicated that we needed only 51 participants (17 in each of three groups) to achieve a power of .95 when alpha was set at .05. Therefore, we targeted a total sample of 60 undergraduate psychology students. However, an extra person signed up to participate, and we ended up with 61 students who participated for class credit. Students were assigned randomly to each of three groups such that the Restudy, Test-No Feedback, and the Test-Feedback conditions contained 20, 21, and 20 participants, respectively.

Materials

Materials were similar to those used by Toppino and Pagano (2020). Lists contained 48 pairs of common English words, half of which had a low intra-pair association value (0.050–0.054) according to the norms published by Nelson, McEvoy, and Schreiber (2004). The remaining pairs had no normative associative connection. Differences in association value were intended to ensure variability in the difficulty of items. For purposes of data analysis, however, we were primarily concerned with participants' perception of item difficulty as reflected by their JOLs.

Procedure

Except for certain details of the instructions and timing, the procedures were the same as those of Toppino and Pagano (2020). During the study phase of the experiment, pairs were presented in an independently determined random order for each participant, and participants studied each pair once or twice. Then, they performed simple math problems for 5 min and took a final cued-recall test on all pairs. The critical events took place during the study phase. After initially studying a pair, participants made a JOL on a scale from 0 to 10, corresponding to their estimate of their probability of recalling the item on a later test. Next, they chose whether to practice the items again sooner (after a short spacing interval filled with two other item presentations) or later (after a long spacing interval in which second occurrences were presented in a new random order after all pairs had been studied at least once). They also could choose to be done with the item (the Done option) in which case they would not encounter it again until the final test. In the Restudy condition, the second occurrence of a pair selected for further practice was a restudy trial in which the full pair was re-presented. In the Test-No

Feedback condition, the second occurrence entailed the presentation of the first word of the pair (cue), and participants tried to recall the second word of the pair (target) before the cue was terminated. The same was true for the Test-Feedback condition except that the full pair was presented as feedback after the cue-word alone was terminated. A 5-s duration was used for the initial study trial in all conditions, the restudy trial in the Restudy condition, the test trial (cue alone) in both Test conditions, and feedback in the Test-Feedback condition.

Participants were instructed thoroughly about the JOL and spacing decisions they had to make on the initial study trial with each pair and practiced this part of the procedure three times. They were further instructed that they should decide to practice again sooner or later with the goal of maximizing their performance on the final test and that they should choose the Done option only when they were highly confident that they knew the pair and would be able to recall it on the final test. In both Test conditions, participants were instructed that, when the practice test occurred, they should try to remember the target word while the cue word was being presented for 5 s, but that they should not say it aloud, type it, or write it; they should try to recall it to themselves. In the Test-Feedback condition, participants also were told that, after trying to recall the target word, the correct answer (i.e., the complete pair) would be presented for 5 s. Although not stressed, it was noted that presenting the correct answer gave participants another opportunity to study the pair.

All pairs were tested in a new random order during the final cued-recall test. The procedure was the same as that for the practice tests in the Test-No Feedback condition except that the cue word was presented alone for 10 s, during which participants tried to type the correct answer.

Data-analysis issues

The data of interest in both Experiment 1 and Experiment 2 concern how perceived item difficulty, measured by JOLs, and the Type of Repeated Practice affected the proportion of trials on which participants chose to practice again sooner and the proportion of trials on which they opted to practice again later.

With respect to JOLs, participants often vary greatly in the absolute level of their JOLs, but our interest was on the effect of relative JOLs within participants. Therefore, we followed a precedent in the literature (e.g., Son, 2004, 2010; Toppino et al., 2018; Toppino & Pagano, 2020) to Vincentize the JOL ratings for each participant. This procedure effectively normalized the JOL data by partitioning each participant's JOLs into three categories, representing items with the lowest, intermediate, and highest JOLs, respectively.

The proportion of trials on which participants chose the Done option is included in the figures illustrating the results but was not included in the data analyses for the following

reasons. First, the purpose of this experiment is to determine the spacing choices participants make when they believe that they need and will benefit from additional practice. The Done option was included so that participants would not be forced to choose a spacing interval to continue practice when they did not believe they needed additional practice in the first place. According to Son (2004), this could lead to a systematically different basis of selection, potentially distorting the data of interest. Second, the proportion of Done choices adds no new information because it is the inverse of the proportion of trials on which participants choose to continue practicing. That is, as the proportion of choices to keep practicing changes (including both Sooner and Later choices), the proportion of Done choices changes in the opposite direction. This also means that there would be complete dependency among the choice alternatives if the proportion of Done choices were included in the data analyses along with the proportion of the Sooner and Later choices.

Lastly, final recall performance is not an essential aspect of the results. Although anticipation of the final recall test was an important part of the rationale presented to participants for studying and practicing, the test itself was not an essential part of the experiment and could have been omitted entirely. It was included only because participants expected a final test and because we wanted to minimize the chances that participants, and potentially future participants, might question the rationale we provided. The most important concern about the final-recall data, however, is that they are fatally confounded by item-selection artifacts, stemming from the fact that participants determined for themselves which items would be

practiced again and after what spacing interval. Nevertheless, the final-recall data are reported in Tables 1 (Experiment 1) and 2 (Experiment 2) of the Appendix in terms of the overall proportions of correct recall for each combination of JOL, Choice, and Type of Repeated Practice. These data should be viewed with extreme caution.

Results

The proportion of participants' choices were analyzed as a function of Practice Condition (Restudy, Test-No Feedback, or Test-Feedback), JOL (Low, Medium, and High) and Spacing Choice (Sooner or Later). The data were submitted to a $3 \times 3 \times 2$ mixed ANOVA with repeated measures on the last two factors. The mean proportions are presented in Fig. 1.

Results revealed a significant main effect of JOL, F (2, 116) = 71.942, MSE = .016, $p < .001$, η^2_p = .554, such that participants became less likely to opt for additional practice (i.e., more likely to choose the Done option) as JOLs increased from low to high, as well as a significant JOL × Practice Condition interaction, F (4, 116) = 2.457, MSE = .016, p = .049, η^2_p = .078, such that participants in the Restudy condition were less likely to choose further practice for the higher JOL items than participants in either of the Test conditions. There also were reliable effects of Choice, F (1, 58) = 10.728, MSE = .277, $p < .002$, η^2_p = .156, and of the JOL × Choice interaction, F (2, 116) = 21.279, MSE = .072, $p < .001$, η^2_p = .268, but these effects were overridden by a significant three-way interaction
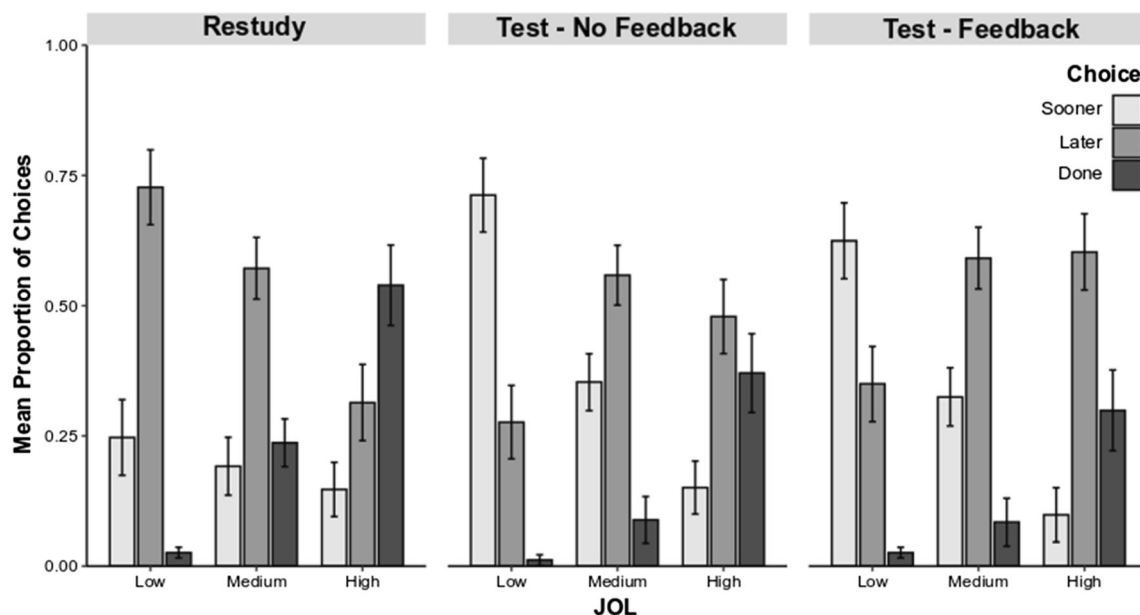

Fig. 1 Mean proportion of trials in Experiment 1 for which the second presentation of a studied item was chosen to occur sooner, later, or not at all ("Done") as a function of Vincentized JOL magnitude ("Low," "Medium," or "High") and practice condition. Error bars represent one standard error of the mean

involving Practice Condition, JOL, and Choice, F (4, 116) = 14.256, MSE = 1.025, $p < .001$, η^2_p = .330.

As can be seen in Fig. 1, participants in the Restudy condition showed a robust overall preference for restudying later rather than sooner, F (1, 19) = 21.775, MSE = .162, $p < .001$, η^2_p = .534, although the difference was especially pronounced for the hard, low-JOL items, F (2, 38) = 3.627, MSE = .071, $p = .036$, η^2_p = .16. In contrast, participants in both of the test conditions chose sooner more often than later for the hardest, low-JOL items although they chose later more frequently for the medium- and high-JOL items. An ANOVA on the data for the Test-No Feedback and Test-Feedback conditions confirmed that the JOL × Choice interaction was significant, F (2, 78) = 46.080, MSE = .073, $p < .001$, η^2_p = .542, while neither the main effect of Practice Condition nor any interaction involving that factor approached significance, all Fs < 1.000.

Discussion

Participants in the Restudy condition exhibited a general preference for restudying after a longer spacing interval although this preference was particularly pronounced for the low-JOL items, which they perceived to be the hardest. This pattern of findings replicates numerous previous studies in which learners chose whether to mass or space restudy opportunities (e.g., Benjamin & Bird, 2006; Toppino et al., 2009; Toppino & Cohen, 2010). In contrast, participants in both of the retrieval practice conditions chose a short spacing interval more often than a long one, with this preference being especially marked for hard, low-JOL items. As JOLs increased, participants chose a short spacing interval less often and a long interval more often, producing a clear preference for a longer spacing interval for the easiest, high-JOL items.

The most important finding was that, whereas there was a marked difference in the pattern of choices made by participants in the Restudy condition and those in the two Test conditions, there was no hint that the pattern of choices made by participants in the Test-Feedback condition differed from that of participants in the Test-No Feedback condition. Successful retrieval practice is essential in the Test-No Feedback condition because learners lose a practice opportunity if retrieval fails on the practice test. However, this is not true in the Test-Feedback condition in which a restudy opportunity in the form of feedback is presented following retrieval practice. Nevertheless, participants in the Test-Feedback condition seemed to ignore the learning potential of feedback and chose spacing intervals as if no feedback would be presented.

These findings replicated the primary results reported by Toppino and Pagano (2020), and showed that the metacognitive error they observed persisted even though participants in the present experiment engaged in covert retrieval practice and in spite of the fact that the importance of feedback

was emphasized by increasing its duration. Thus, we obtained no evidence for the hypothesis that learners ignore the benefits of anticipated feedback when making strategy choices because they succumb to implicit demands to look good by performing well on practice tests regardless of feedback. Similarly, the result fails to support the hypothesis that participants have discounted the learning potential of feedback in past experiments because of its comparatively brief presentation duration in those studies.

Choosing a short spacing interval for hard items when practice tests are followed by corrective feedback is a metacognitive error to the extent that choosing a longer spacing interval would lead to better recall. That this choice pattern constituted a metacognitive error was demonstrated by Toppino and Pagano (2020, Experiment 2), using an honor/dishonor paradigm in which some spacing choices for each participant were honored while other choices were dishonored, with the opposite spacing being substituted instead. They found that honoring the choice of a short spacing interval benefited recall in the absence of post-test feedback, whereas a longer spacing interval led to better recall regardless of participants' choices when practice tests were followed by feedback. The present experiment did not use the honor/dishonor paradigm, and the recall data must be viewed with great caution, as explained earlier, but trends in the recall data are consistent with the metacognitive-error interpretation. The recall advantage of longer over shorter spacing intervals was greater in the Test-Feedback than in the Test-No Feedback condition, and recall was uniformly better in the Test-Feedback condition than in the Test-No Feedback condition when a longer spacing interval was chosen (Table 1).

As a final point, there was one minor difference between the present results and those obtained by Toppino and Pagano (2020), which was inconsequential for the purposes of the current research. Participants in Toppino and Pagano's experiment preferred a short spacing interval for items with medium JOLs, whereas learners in the present study preferred a long spacing interval for these items. This discrepancy may reflect sampling differences or may be the effect of procedural differences between the experiments. More generally, however, it may not be surprising that some variability in results was obtained in the condition involving medium JOL items for which learners' preference for short or long spacing intervals tends to be smallest.

Experiment 2

In discussing the results of our first experiment, our emphasis was on the fact that retrieval practice yielded similar results

regardless of whether or not it was followed by feedback. It also may be instructive to compare the Test-Feedback condition with the Restudy condition. The full pair was presented on both the initial study trial and on the additional practice trial in both of these conditions, which differed only in that learners in the Test-Feedback condition took a practice test immediately before the pair's second presentation. This difference, however, produced a dramatic difference in spacing choices, particularly for the hardest items. In fact, when a practice test was expected prior to the second presentation of the full pair, participants, who definitely knew the full pair was going to be presented, behaved as though it would not be presented at all. The question is why the effect of expecting a presentation of the full pair on a second practice trial is nullified when learners expect it to be preceded by a practice test.

One possibility is that learners interpret the presentation of the full pair to be a qualitatively different kind of event when it is expected to occur by itself as a restudy trial and when it is expected to follow a prior practice test as feedback. Learners report using retrieval practice as part of their real-life learning regimen primarily to diagnose their level of learning rather than to improve their learning (e.g., Hartwig & Dunlosky, 2012; Kornell & Bjork, 2007). If feedback is viewed as an integral part of the diagnostic function of testing, learners may fail to consider it as the study opportunity it is. If that is the case, we might find that learners' pattern of choices would change if the presentation of the full pair following a practice test were framed as a post-test study opportunity rather than as feedback.

The present experiment incorporated three Practice conditions: A Restudy condition that was a replication of the comparable condition in Experiment 1, and two conditions in which retrieval practice was followed by feedback. The latter conditions differed only with respect to instructions that were intended to frame the presentation of the full pair after retrieval practice as feedback as in Experiment 1 (Test-Feedback condition) or as a post-test study opportunity (Test-Feedback Framed). To the extent that learners typically fail to construe practice-test feedback as a learning opportunity, we should find the previously obtained pattern of spacing choices in the Test-Feedback condition. However, the pattern of spacing choices in the Test-Feedback-Framed condition should shift away from that of the Test-Feedback condition and toward the pattern obtained in the Restudy condition.

Method

Participants

Participants were 62 undergraduate psychology students assigned randomly to three Practice Conditions, resulting in

21, 20, and 21 participants in the Restudy, Test-Feedback, and Test-Feedback-Framed Conditions, respectively.

Materials, procedures, and data analysis

Materials and procedures were identical to those in Experiment 1 with the following exceptions. There was no Test-No Feedback condition. The Test-Feedback and Test-Feedback-Framed conditions were identical, except for the instructions. As in Experiment 1, the instructions in the Test-Feedback condition referred to additional practice as "practice tests." These instructions allowed the presentation of the full pair following the tests to be interpreted as feedback, although, as detailed below, it was explicitly mentioned that this provided another opportunity to study the pair. In the Test-Feedback-Framed condition, the instructions referred to additional practice as "restudy opportunities." Thus, a practice test followed by feedback was framed as a restudy opportunity preceded by a practice test.

Early in the instructions, the Test-Feedback group received the following statement as part of their instructions:

> The pairs will be presented one at a time for 5 seconds, and you should try to learn each pair so that you can remember it on a final test. You also will have the opportunity to take a practice test – if you choose to do so – on each pair prior to the final test.

The corresponding portion of the instructions for the Test-Feedback-Framed group was:

> The pairs will be presented one at a time for 5 seconds, and you should try to learn each pair so that you can remember it on a final test. You also will have the opportunity to study the pair again – if you choose to do so – prior to the final test. The opportunity to re-study a pair will begin with a chance to test yourself on the pair first.

Other differences between the instructions involved the substitution of terminology (e.g., "restudy" or "restudy opportunity" for "test" or "practice test") along with occasional minor grammatical changes to accommodate the terminological differences. In the sample below, only the italicized words were changed from one group to the other in a paragraph describing the details of the practice test. The segment includes the fact that re-presenting the full pair after the test would allow another opportunity to study the pair. The exact paragraph for participants in the Test-Feedback condition was the following:

> During the *practice test*, the top word will be presented alone for 5 seconds. You should try to remember the bottom word, but do not say it aloud,

type it, or write it. Just try to recall it to yourself. After 5 seconds, the correct answer (i.e., the complete pair) will appear for 5 seconds, giving you another opportunity to study the pair.

Participants in the Test-Feedback-Framed condition received the identical paragraph, except that "practice test" was changed to "restudy opportunity."

Finally, treatment of JOLs and Done responses was the same as in Experiment 1. Also, for the same reasons as in Experiment 1, the final recall data do not support clear conclusions but are presented in the Appendix for the sake of completeness.

Results and discussion

The proportion of participants' choices were analyzed as a function of Practice Condition (Restudy, Test- Feedback, or Test-Feedback-Framed), JOL (Low, Medium, and High) and Spacing Choice (Sooner or Later). The data were submitted to a $3 \times 3 \times 2$ mixed ANOVA with repeated measures on the last two factors. The mean proportions are presented in Fig. 2.

The results yielded a significant main effect of JOL, F (2, 118) = 32.311, $MSE = .030$, $p < .001$, $\eta^2_p = .354$, such that participants chose to continue practice less (i.e., chose the Done option more) as JOLs increased from low to high. Other significant effects included a main effect of Choice, F (1, 59) = 25.739, $MSE = .226$, $p < .001$, $\eta^2_p = .304$, the Choice × Practice Condition interaction, F (2, 59) = 6.589, $MSE = .226$, $p < .003$, $\eta^2_p = .183$, and the JOL × Choice interaction, F (2, 118) = 3.509, $MSE = .105$, $p = .033$, $\eta^2_p = .056$. However, these effects were qualified by a significant three-way interaction (Practice Condition × JOL × Choice), F (4, 118) = 4.744, $MSE = .105$, $p < .002$, $\eta^2_p = .139$.

The three-way interaction was probed by separate ANOVAs on the data of each Practice Condition. Participants in the Restudy condition exhibited a strong overall preference for restudying later as opposed to sooner, F (1, 20) = 34.702, $MSE = .167$, $p < 001$, $\eta^2_p = .634$, replicating the major finding of Experiment 1 when additional practice involved restudying. The JOL × Choice interaction, however, was not reliable, F (2, 40) = 0.561, $MSE = .121$, $\eta^2_p = .027$. In the Test-Feedback condition, participants preferred to take a practice test sooner for hard, low-JOL items but later for easier, higher-JOL items. This was confirmed by a significant JOL × Choice interaction, F (2, 38) = 9.108, $MSE = .136$, $p < .001$, $\eta^2_p = .324$, replicating the major finding of Experiment 1 when additional practice involved retrieval practice. In the Test-Feedback-Framed condition, there was a substantial general preference for participants to take a practice test later rather than sooner, F (1, 20) = 11.024, $MSE = .289$, $p < .004$, $\eta^2_p = .355$, and the JOL × Choice interaction was not significant,
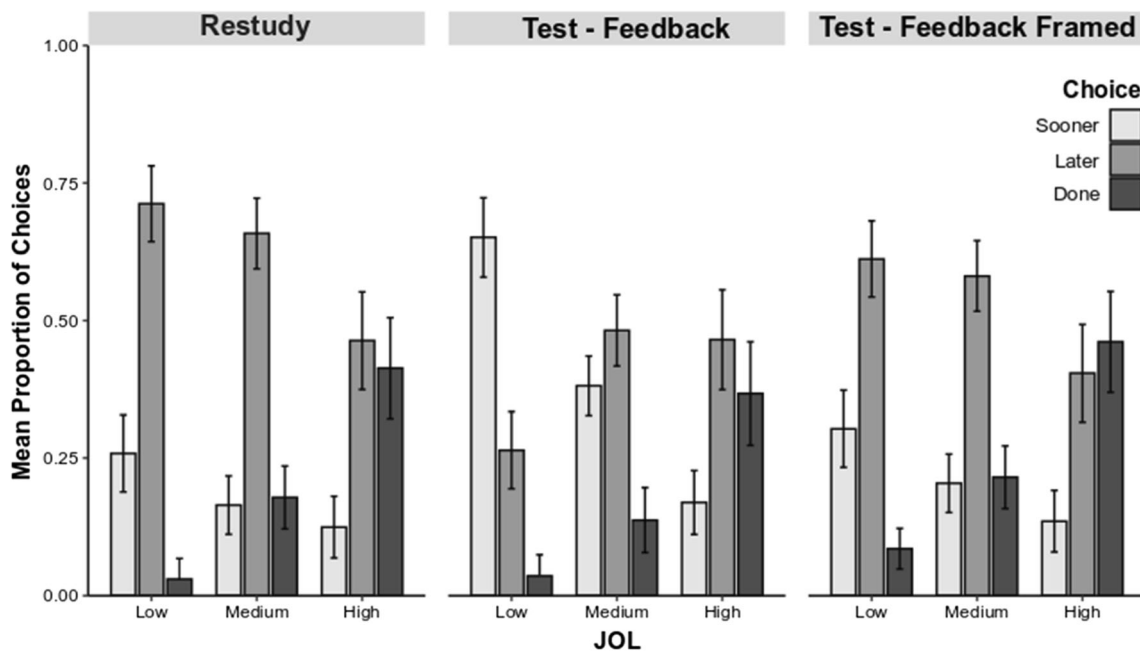

Fig. 2 Mean proportion of trials in Experiment 2 for which the second presentation of a studied item was chosen to occur sooner, later, or not at all ("Done") as a function of Vincentized JOL magnitude ("Low," "Medium," or "High") and practice condition. Error bars represent one standard error of the mean

$F (2, 40) = 0.539$, $MSE = .058$, $\eta^2_p = .026$. This pattern of choices is similar to that exhibited by participants in the Restudy condition and is markedly different from the pattern displayed by the Test-Feedback condition. Thus, framing feedback as a restudy opportunity led participants to alter their preferences for whether it would be best to take a practice test after a long or a short spacing interval. This suggests that, without intervention, participants do not typically construe feedback to be a restudy opportunity.

General discussion

In associative learning, both learning and retention are facilitated when practice tests are followed by feedback involving presentation of both the cue and the target items (e.g., Pashler et al., 2005). Although longer spacing intervals impair performance on practice tests, the presentation of feedback allows learners to reap substantial benefits from these longer spacing intervals (e.g., Pashler et al., 2003; Toppino & Pagano, 2020). Despite the potent effect of feedback on learning and retention, however, learners often seem to discount or even ignore the beneficial effect of feedback when making metacognitive decisions. After receiving feedback on a test, learners may fail to predict that their future memory will be improved or that they may need less additional practice than they would have needed if feedback had not been presented (Kornell & Rhodes, 2013). Even when they predict better memory after receiving feedback, they greatly underestimate the magnitude of its effect (Sitzman et al., 2016). In deciding whether their future practice will involve restudying or taking practice tests, learners' tendency to choose retrieval practice may be no greater when they expect it to be followed by feedback than when they expect no feedback, especially when they have no prior practice on the task (e.g., Kornell & Son, 2009; Toppino et al., 2018; Tullis et al., 2018). Also, learners' choices of whether retrieval practice should occur after a short or a long spacing interval may fail to be influenced by their expectation of the presence or absence of feedback after practice tests, even though longer spacing intervals would seem to be advantageous when feedback is provided.

In the present experiments, we sought to better understand why learners' metacognitive decisions are so little influenced by feedback even though it is so demonstrably effective in facilitating learning. We did this by adopting the paradigm used by Toppino and Pagano (2020) and examining the effect of factors that might be revealing

of why the expectation of feedback may fail to impact learners' metacognitive decisions.

In Experiment 1, we considered whether learners discount the effect of feedback because its duration is brief or because of implicit demands that pressure learners to prioritize retrieval-practice performance, despite being instructed to make decisions that will benefit final recall, not practice-test recall. First, we equated the duration of all learning events so that, for example, the full word pair was presented for a duration of 5 s when it was presented as the initial study opportunity, as a restudy opportunity, or as feedback following a practice test. Second, we required that retrieval practice be completed covertly rather than overtly so that learners could gain the benefit of retrieval practice while avoiding any unrelated pressure to perform well on practice tests for the sake of appearance. However, our results provided no evidence that either the duration of feedback or the demands of overt retrieval practice were complicit in learners' tendency to ignore the beneficial effect of feedback when making metacognitive choices during learning. That is, the presence or absence of feedback produced no hint of a difference in the spacing choices of learners who engaged in retrieval practice, replicating the essential aspects of Toppino and Pagano's (2020) findings.

In Experiment 2, we considered the possibility that learners simply do not construe feedback after retrieval practice to be a learning opportunity. It is well established that, when learners opt to engage in retrieval practice while learning, they do it primarily to diagnose their level of learning (e.g., Hartwig & Dunlosky, 2012; Kornell & Bjork, 2007; Wissman et al., 2012). Although retrieval itself may convey diagnostic information, feedback provides much more definitive information. When learners anticipate that retrieval practice will be followed by feedback, their focus may be on its role in diagnosing learning. Thus, even though learners may know that re-presenting both members of a pair provides a learning opportunity, they may fail to access that information in contexts that emphasize feedback's diagnostic function. If that is the case, the influence of feedback on learners' metacognitive decisions might be altered by instructions designed to highlight and increase access to the learning potential of feedback. Therefore, we presented feedback in the usual way (Test-Feedback condition), or we framed it in the instructions as a restudy opportunity preceded by a practice test (Test-Feedback Framed). When the presentation of the full word pair after the practice test was framed as feedback in the usual way, learners manifested a

preference for a short spacing interval for the hardest items, a strategy we had seen in Experiment 1 for both testing conditions, regardless of whether the test was or was not followed by feedback. However, when feedback was framed as a restudy opportunity, learners' choice behavior was altered dramatically. Under these circumstances, they preferred a long spacing interval regardless of JOL, a strategy that learners use when additional practice involves only a restudy opportunity (i.e., no retrieval practice).

It is important to note that learners could have exhibited the same choice behavior when presentation of the full pair was framed as feedback in the Test-Feedback condition if they had been cognizant of its potential as a study opportunity. However, they did not. This fact suggests that learners in the Test-Feedback condition did not construe feedback to provide a restudy opportunity as they made their choices about spacing intervals. The question is why.

Our interpretation is that the context in which the intact pair is presented as feedback determines whether learners will more readily access its diagnostic function or its restudy function. When the intact pair is presented for initial study, learners seem to have no trouble interpreting it as a study opportunity. When intact pairs are presented as feedback, however, they occur in the context of practice tests, which learners commonly view as a means to diagnose learning (e.g., Hartwig & Dunlosky, 2012; Kornell & Bjork, 2007). Perhaps it should not be surprising that learners fail to access the restudy function of feedback under these circumstances. When feedback is framed as a restudy opportunity preceded by a practice test, however, the restudy function of feedback is likely to be accessed with consequent changes in behavior. Thus, we found that framing feedback as a restudy opportunity affects learners' metacognitive decisions about whether a long or short spacing interval should separate the initial study opportunity from a subsequent practice test. We also would expect an effect of framing in other situations in which the anticipation of feedback might affect metacognitive choices. For example, if learners were given the choice between restudying or taking a practice test as has been done in some previous research (e.g., Toppino et al., 2018; Tullis et al., 2018), we would expect framing feedback as a restudy opportunity to increase the frequency with which learners prefer to engage in retrieval practice.

Given the powerful effects that feedback has on learning and retention, particularly when retrieval practice is unsuccessful (e.g., Pashler et al., 2003; Pashler et al., 2005; Pashler et al., 2007; Toppino & Pagano, 2020), it is striking that learners so often fail to exhibit an appreciation of the learning potential of feedback in studies of metacognition (e.g., Kornell & Rhodes, 2013; Toppino et al., 2018; Toppino & Pagano, 2020; Tullis et al., 2018). Although this lack of appreciation is consistent with the aforementioned survey data in which learners report using retrieval practice to diagnose rather than to enhance learning (e.g., Hartwig & Dunlosky, 2012; Kornell & Bjork, 2007), it should be noted that the metacognitive failings related to feedback have been observed in experimental tasks with which learners may have had little experience. They may be unlikely to access and weigh the learning potential of feedback under these circumstances. They may be more likely to draw upon previous experience and to spontaneously access the restudy function of feedback in more familiar learning settings. As an initial step, future research might question whether metacognitive awareness of the learning benefits of feedback are more accessible with ecologically relevant learning materials (e.g., factual material; Kang et al., 2011; Smith & Kimball, 2010) or familiar tasks (e.g., flashcards; Kornell, 2009). Even with unfamiliar tasks, learners might gain a greater appreciation of feedback as a restudy opportunity without explicit instructions if they were to receive multiple learning experiences, each of which entails engaging in retrieval practice with feedback.

Our results make it clear that framing feedback as a restudy opportunity markedly alters learners' metacognitive choices such that they are much more likely to schedule a practice test with feedback after a longer rather than a shorter spacing interval. What is less clear is whether this constitutes an efficacious change in strategy. In favor are the facts that the benefits of spaced practice are well known (e.g., Toppino & Gerbier, 2014), and providing feedback following a practice test allows learners to benefit from longer spacing intervals even if retrieval practice is unsuccessful. Furthermore, feedback may enable test-potentiated learning in which studying is more effective following a retrieval attempt (e.g., Arnold & McDermott, 2013). It also is possible that participants learn more from feedback when it is identified as a restudy opportunity, although this remains an open question. The efficacy of the strategy, however, is not a foregone conclusion. For example, retrieval practice is well established as a potent learning activity (e.g., Roediger & Karpicke, 2006a, 2006b). Framing feedback as a restudy opportunity could focus learners' attention on the restudy

opportunity while weakening the beneficial effect of retrieval practice by reducing the attention and effort they devote to the practice test per se. The net effect might be that final recall is not better, or possibly even worse, when feedback is framed as a restudy opportunity. Unfortunately, the recall data from Experiment 2 provide no hints in the form of clear suggestive trends, and the issue needs to be addressed in future research.

Appendix

As noted in the *Method* section of Experiment 1, the recall data was, by design, seriously confounded by participants' spacing choices and must be viewed with extreme caution. Nevertheless, we have included the overall proportions of correct recall for each combination of variables in Tables 1 and 2 below.

Table 1 Proportions of correct final cued recall in Experiment 1 collapsed over participants

Practice condition	Low JOLs			Medium JOLs			High JOLs		
	Sooner	Later	Done	Sooner	Later	Done	Sooner	Later	Done
Restudy	.408	.529	.625	.754	.694	.583	.805	.714	.665
Test-No FB	.385	.423	.250	.675	.665	.704	.865	.778	.817
Test-FB	.416	.602	.250	.600	.727	.538	.833	.888	.690

Table 2 Proportions of correct final cued recall in Experiment 2 collapsed over participants

Practice condition	Low JOLs			Medium JOLs			High JOLs		
	Sooner	Later	Done	Sooner	Later	Done	Sooner	Later	Done
Restudy	.683	.768	.429	.516	.704	.612	.646	.763	.682
Test-FB	.654	.699	.615	.672	.778	.756	.778	.806	.717
Test-FB Framed	.604	.679	.375	.765	.686	.397	.803	.769	.515

References

Arnold, K. M., & McDermott, K. B. (2013). Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 940–945. https://doi.org/10.1037/a0029199

Benjamin, A. S., & Bird, R. D. (2006). Metacognitive control of the spadcing of study repetitions. *Journal of Memory and Language, 55*(1), 126-137.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, *19*(1), 126–134. https://doi.org/10.3758/s13423-011-0181-y

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports, 18*(3), 879-919.

Kang, S. H. K., 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*(4–5), 528–558. https://doi.org/10.1080/09541440601056620

Kang, S. H. K., Pashler, H. Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology, 103*, 48-59.

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General*, *138*(4), 469–486. https://doi.org/10.1037/a0017341

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory*, *17*(4), 471–479. https://doi.org/10.1080/09658210802647009

Karpicke, J. D., & Roediger, H. L. (2008). The Critical Importance of Retrieval for Learning. *Science*, *319*(5865), 966–968. https://doi.org/10.1126/science.1152408

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology, 23*, 1297-1317

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*(2), 219–224. https://doi.org/10.3758/BF03194055

Kornell, N., & Rhodes, M. G. (2013). Feedback reduces the metacognitive benefit of tests. *Journal of Experimental Psychology: Applied, 19*(1), 1–13. https://doi.org/10.1037/a0032147

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17*(5), 493–501. https://doi.org/10.1080/09658210902832915

Kuhbander, C. & Emmerdinger, K. J. (2019). Do students really prefer repeated rereading over testing when studying textbooks? A reexamination. *Memory, 27*(7), 952-961.

Morehead, K., Rhodes, M. G., & DeLozier, S. (2016). Instructor and student knowledge of study strategies. *Memory, 24*(2), 257-271.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In K.W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 26, pp. 125-141). Academic Press.

Nelson, D. L., McEvoy, C. L, & Schreier, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402–407. https://doi.org/10.3758/BF03195588

Pashler, H., Cepeda, N. J., Wixted, J. T., and Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(1), 3-8.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin & Review, 14*(2), 187-193.

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

Pyc, M. A., & Dunlosky, J. (2010). Toward an understanding of students' allocation of study time: Why do they decide to mass or space their practice? *Memory & Cognition, 38*(4), 431-440.

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

Roediger, H. L., & Karpicke, J. D. (2006a). The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science, 1*(3), 181–210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., & Karpicke, J. D. (2006b). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

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

Sitzman, D. M., Rhodes, M. G., & Kornell, N. (2016). The influence of feedback on predictions of future memory performance. *Memory & Cognition, 44*, 1102-1113.

Smith, M. A., Roediger, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(6), 1712-1725.

Smith, T. A. & Kimball, D. R. (2010). Learning from feedback: Spacing and the delay-retention effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 80-95.

Soderstrom, N. C., & Bjork, R. A. (2014). Testing facilitates the regulation of subsequent study time. *Journal of Memory and Language, 73*, 99-115

Son, L. K. (2004). Spacing One's Study: Evidence for a Metacognitive Control Strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 601–604. https://doi.org/10.1037/0278-7393.30.3.601

Son, L. K. (2010). Metacognitive control and the spacing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 255–262. https://doi.org/10.1037/a0017892

Storm, B. C., Friedman, M. C., Murayama, K., & Bjork, R. A. (2014). On the transfer of prior tests or study events to subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(1), 115–124. https://doi.org/10.1037/a0034252

Tauber, S. K., Witherby, A. E., Dunlosky, J., Rawson, K. A., Putnam, A. L., & Roediger, H. L. (2018). Does covert retrieval benefit earning of key-term definitions? *Journal of Applied Research in Memory and Cognition, 7*, 106-115.

Toppino, T. C., & Cohen, M. S. (2010). Metacognitive control and spaced practice: Clarifying what people do and why. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(6), 1480–1491. https://doi.org/10.1037/a0020949

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 60, pp. 113–189). Burlington: Academic Press.

Toppino, T. C., & Pagano, M. J. (2020). Metacognitive control over the distribution of retrieval practice with and without feedback and the efficacy of learners' spacing choices. Memory & Cognition. Advance online publication. https://doi.org/10.3758/s13421-020-01100-x

Toppino, T. C., Cohen, M. S., Davis, M. L., & Moors, A. C. (2009). Metacognitive control over the distribution of practice: When is spacing preferred? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1352–1358. https://doi.org/10.1037/a0016371

Toppino, T. C., LaVan, M. H., & Iaconelli, R. T. (2018). Metacognitive control in self-regulated learning: Conditions affecting the choice of restudying versus retrieval practice. *Memory & Cognition, 46*(7), 1164–1177. https://doi.org/10.3758/s13421-018-0828-2

Tullis, J. G., Feichter, J. L., & Benjamin, A. S. (2018). The efficacy of learners' testing choices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(4), 540–552. https://doi.org/10.1037/xlm0000473

Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2012). How and when do students use flashcards? *Memory, 20*, 568–579. https://doi.org/10.1080/09658211.2012.687052

Yan, V. X., Thai, K.-P., & Bjork, R. A. (2014). Habits and beliefs that guide self-regulated learning: Do they vary with mindset? *Journal of Applied Research in Memory and Cognition, 3*(3), 140–152. https://doi.org/10.1016/j.jarmac.2014.04.003

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.