

The Benefits of Successive Relearning on Multiple Learning Outcomes

Philip A. Higham¹, Bettina Zengel², Laura Bartlett³ and Julie A. Hadwin⁴

¹University of Southampton

²University of Essex


³London School of Economics


⁴Liverpool Hope University


© 2021, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/edu0000693

Author Note

Philip A. Higham  <https://orcid.org/0000-0001-6087-7224>

Bettina Zengel  <https://orcid.org/0000-0002-0871-3158>

Laura Bartlett  <https://orcid.org/0000-0001-5202-4504>

Julie A. Hadwin  <https://orcid.org/0000-0002-1776-7940>

We have no known conflict of interest to disclose.

This work was supported in part by a grant from the Association for Psychological Science (APS) Fund for Teaching and Public Understanding of Psychological Science awarded to Philip A. Higham and Julie A. Hadwin.

Supplemental materials are available at:

https://osf.io/nx4va/?view_only=8fc6db2c67d545f784463ad4356bbc08

Correspondence concerning this article should be addressed to Philip A. Higham, School of Psychology, University of Southampton, Highfield, Southampton, UK, SO17 1BJ.

E-mail: higham@soton.ac.uk

Abstract

Successive relearning involves repeated retrieval practice of the same information (with feedback) over multiple, spaced sessions. We implemented successive relearning in an introductory psychology class to explore potential learning benefits. After each weekly lecture, students were sent links via email to engage in three learning practice sessions, each separated by two days. Half the students engaged in successive relearning (*relearn* condition), answering 20 fill-in-the-blank questions with corrective feedback. Within each session, correctly answered questions were dropped, whereas incorrectly answered questions were presented up to 2 more times. The other half of students restudied the same 20 sentences without blanks twice per session (*restudy* condition). Unlike previous research, we controlled the exposure duration of the learning materials between the relearn and restudy conditions. Learning practice sessions continued throughout the remaining 10 weeks of the semester, with students alternating each week between the relearning and restudying tasks. Recall of course material at the end of the semester was better for relearning compared to restudying. Increased recall during relearning sessions was associated with further learning benefits including improved metacognition, increased self-reported sense of mastery, increased attentional control, and reduced anxiety. Individual differences were not associated with the benefit of relearning over restudying in the retention tests. Qualitative feedback indicated that students found successive relearning to be enjoyable and valuable. Our research indicates that successive relearning is a valuable addition to any university course and is easy to implement using digital resources.

Key words: Successive relearning, spacing effect, distributed practice, retrieval practice, individual differences

Educational Impact and Implications Statement

Successive relearning is a highly effective learning technique that involves students taking several practice tests on to-be-learned material (with corrective feedback), with each test separated by a time interval of a few days. We had students engage in successive relearning in a large introductory Psychology course throughout a semester and found that they not only remembered more course content (compared to students who restudied the material for approximately the same amount of time), but they were also less anxious, felt that they had better attentional control and more mastery over the material, and had more accurate metacognitions. These benefits were not observed immediately but required several practice testing sessions before they were revealed. Overall, the results suggest that there are multiple learning outcomes that benefit from repeated testing sessions that are spaced apart over time and that sessions like these should be implemented regularly in real classrooms.

The Benefits of Successive Relearning on Multiple Learning Outcomes

In a review of ten techniques students use to enhance their learning, Dunlosky et al. (2013) identified *spaced practice* and *retrieval practice* as having the highest utility. Regarding spaced practice, several studies have shown that repeated practice taking place over multiple, short, temporally-spaced learning sessions (*spaced practice*) yields better and more enduring memory than repeated practice occurring in a single, long learning session (*massed practice*), a difference referred to as the *spacing effect* (review by Benjamin & Tullis, 2010). Regarding retrieval practice, over a century of research has demonstrated that students who practice retrieving information have better and longer memory retention for that information compared to students who engage with the material more passively (e.g., restudying). This memory advantage for retrieved versus restudied material is referred to as the *testing effect* (review by Rowland, 2014). In addition to these direct benefits on memory, retrieval practice can also produce other learning benefits, particularly if it is repeated. For example, Healy et al. (2017) found that lectures interspersed with brief periods of retrieval practice yielded high levels of student-reported cognitive engagement during lectures. Additionally, Szpunar et al. (2013) found that interpolated testing (brief testing episodes interspersed throughout a learning event) during lectures led to low levels of mind-wandering, as well as low anxiety and low negative affect towards the final cumulative exam. These positive effects on attention and anxiety led to better learning of the final lecture segment, even though the material was new to students, a finding referred to as the *forward testing effect* (review by Yang et al., 2018).

Successive Relearning

The current research investigated a third learning technique, not included in Dunlosky et al.'s (2013) review, that combines repeated retrieval practice and spaced practice over long intervals into a single technique called *successive relearning* (Baird, 1979). Successive

relearning requires that students complete a memory test for taught material with corrective feedback until they *master* the material. Mastery involves correctly recalling all answers to a specified *criterion level*, which can be one correct recall or more. Items that reach criterion are eliminated from further testing, whereas incorrectly answered questions are presented to be answered again later in the same testing session (*dropout methodology*). The number of recall attempts to reach criterion is known as *trials-to-criterion* and the session ends when all questions are recalled to criterion. This whole process is repeated with the same questions across a number of spaced relearning sessions.¹

A small evidence base suggests that successive relearning produces excellent and durable memory for learned material – more than either single-session testing or spaced practice on its own (e.g., (Bahrick, 1979; Bahrick et al., 1993; Bahrick & Hall, 2005; Janes et al., 2020; Rawson et al., 2013; Rawson & Dunlosky, 2011, 2012, 2013; Vaughn et al., 2016). For example, Rawson, Vaughn, Walsh, & Dunlosky (2018) found that participants recalled 20% of learned information one week after a single learning session. However, if the same information was successively relearned three times (with mastery) over spaced intervals, recall after one week increased to 80% (Cohen's *d* for the difference = 4.19). Furthermore, three weeks after the fourth relearning session, participants still recalled 77% of the material, suggesting that successive relearning produced lasting benefits.

Confounding with Exposure

One potential problem with existing studies is that the effect of successive relearning is typically confounded with exposure to the to-be-learned material. For example, Rawson et al.'s (2018) finding of better retention of learned material after three versus one successive

¹ Successive relearning is very similar to spaced retrieval practice, which is another highly effective learning technique (e.g., Karpicke & Bauernschmidt, 2011; Lyle et al., 2020). The main difference between the two techniques is that successive relearning involves mastery, dropout methodology, and trials-to-criterion whereas spaced retrieval practice does not necessarily incorporate these methodological features. Hence, successive relearning might be viewed as a more efficient form of spaced retrieval practice, where the focus is on students mastering material that they have not yet learned.

relearning sessions might simply reflect differential exposure to the material; other elements of successive relearning (i.e., retrieval practice, spacing, feedback, and/or mastery) may not be necessary to achieve this benefit. The confound of exposure may also have more subtle effects. For example, Rawson and Dunlosky (2013; Experiment 1) asked participants to read a passage in an initial learning session and then answer questions about it. Incorrectly answered questions were presented again in the same session until mastery was achieved. The number of intervening items (lag) between repetitions was varied within the initial session (lag 0 [massed], 1, 3 or 7 [different levels of spacing]). Consistent with previous work on the spacing effect, greater lags during initial learning led to facilitated recall performance in the first relearning session on Day 3, but this advantage was attenuated in a subsequent relearning session on Day 8.² By the final test on Day 10, the initial spacing effect was completely eliminated, an effect dubbed *relearning override* (Vaughn et al., 2016).

Some researchers have argued that relearning override effects are of theoretical importance as they are consistent with some frameworks (e.g., two-stage framework; Kornell et al., 2015) and inconsistent with others (e.g., retrieval effort hypothesis; Pyc & Rawson, 2009). However, the timing data from some of these studies indicates that relearning override could be potentially explained at least in part by the exposure confound. For example, in the Rawson and Dunlosky (2013) study described earlier: (1) learners spent more time learning the spaced versus the massed items to criterion in the initial session (because of more between-repetition forgetting in the spaced conditions), (2) in the first successive relearning session, more time was needed for massed items to be mastered to criterion because of forgetting since the initial session, reducing the discrepancy in both the amount of learning

² Because each relearning session begins with an initial attempt at answering all questions followed by feedback, each session acts as an interim memory test.

and total time-on-task between different lag conditions and (3) a reduction in the discrepancy between the learning level and total time-on-task between the lag conditions continued in the third successive relearning session. By the final test, the advantage of initially spaced items over initially massed ones was eliminated and the total time-on-task was equated.

Interestingly, in their third experiment, there was a residual initial spacing effect on the final test, but there was also a residual time-on-task discrepancy favoring spaced items. As long as total exposure time over all sessions was equated, complete relearning override occurred. If exposure time was not equated, then relearning override was incomplete and a residual spacing effect was observed. This analysis identifies the need for better control over spacing and exposure to the learned material.

In contrast to successive relearning research, where exposure time is typically not controlled, many single-session testing-effect experiments include a restudy control condition. In this condition, participants restudy (rather than practice retrieving) material, thereby equating exposure time between the tested and non-tested conditions. Rowland (2014) argued that failure to equate exposure time with a restudy control "...can potentially produce an upwardly biased estimate of the retention benefit that results from the act of retrieval itself" (p. 1433). One reason that research on successive relearning has not adopted a similar design may be that relearning time is dependent on participants' performance (e.g., slow learners will spend more time per item than fast learners) rather than controlled by the experimenter. One way to overcome this problem is to provide participants the opportunity to self-regulate both relearning and restudy times (e.g., Rawson et al., 2013). However, self-regulated decisions about how long to restudy versus relearn items are unlikely to be equivalent, so this approach only partially addresses the problem. Some researchers have considered the exposure confound in successive relearning studies to be so serious that the

data have been deemed unsuitable for meta-analyses of spaced practice (e.g., Cepeda et al., 2006).

Metacognition and The Learning Experience

Several studies have found that the accuracy of metacognitive processes (e.g., ability to predict later memory performance) is correlated with academic achievement (see review by Ohtani & Hisasaka, 2018). Prior research, however, paints a bleak picture regarding the metacognitive accuracy of both spaced practice and retrieval practice. Despite both techniques being highly efficacious for memory, learners do not seem to recognize the benefits. For example, Baddeley and Longman (1978) found that their participants considered a massed (vs. spaced) practice schedule to be more satisfactory for their prior training, despite clear evidence that a spaced practice schedule produced better learning. Their participants also preferred (and were keener to engage in) a massed practice schedule for future training. Relatedly, other researchers have found that participants erroneously predict either no difference or better performance following a massed (vs. spaced) practice schedule (e.g., Emeny et al., 2021; Logan et al., 2012; Simon & Bjork, 2001; Zechmeister & Shaughnessy, 1980). In terms of the testing effect, when asked to make item-based judgments of learning, individuals often indicate that they will remember restudied material better than tested material, when the opposite is nearly always true (e.g., Karpicke, 2009; Kornell & Son, 2009; Potts & Shanks, 2014; Roediger & Karpicke, 2006). However, further evidence suggests that if individuals experience the learning benefits firsthand and these are made clear and obvious, learners come to appreciate the efficacy of both spaced practice and retrieval practice (improving their metacognitive accuracy as a result; e.g., Logan et al., 2012; Tullis et al., 2013). Because successive relearning involves repeated sessions with feedback, it is conceivable that it provides the kind of experience necessary to improve

metacognitive accuracy. Such improvement may be important to motivate learners to independently engage with successive relearning in other teaching and learning contexts.

Successive relearning has the potential to have positive effects on other aspects of the student learning experience. Previous research has found, for example, that repeated testing (e.g., interpolated testing during lectures) was associated with several positive student learning outcomes such as reduced mind wandering, enhanced cognitive engagement, and lower reports of anxious feelings and negative affect towards a final test (Healy et al., 2017; Szpunar, 2017; Szpunar et al., 2013; see also Agarwal et al., 2014). More recently, Yang et al. (2020) found that low-stake repeated testing enhanced new learning and the forward testing benefits were observed regardless of learners' level of test anxiety. While a small literature is emerging on the benefits of testing for the learning experience, no previous study has explored whether the process of successive relearning has a similar broad positive impact.

Overview of The Current Study

This study investigated the effect of successive relearning on the acquisition and retention of taught material in a large undergraduate psychology module. After each weekly lecture, students practiced learning the lecture content two, four, and six days later. There were two types of learning practice, successive relearning and restudying, and each student engaged in one type of practice in all three practice sessions each week. For successive relearning, students answered fill-in-the-blank questions about that week's lecture. The session involved mastery learning (criterion level = 1) with feedback and dropout methodology, as is typical in successive relearning research. For restudying, students read statements comparable to the fill-in-the-blank questions used for successive relearning (i.e., the blanks were filled in and there was no opportunity for students to enter any text). Students restudied/relearned the same material in all three sessions per week. In the following week,

new questions/statements were introduced to accommodate the new lecture content. At the end of the semester, all students completed two fill-in-the-blank criterial tests. The first versus second test was administered, on average, 42 days versus 68 days following the lecture during which the tested material was learned, respectively.

To ensure no student was disadvantaged relative to any other, the students switched learning practice tasks each week; that is, across 10 weeks, each student completed five weeks of relearning and five weeks of restudying, alternating each week. To control the exposure to learning practice material, students had a limited time to answer the questions (and read the feedback) in the relearn condition such that the average exposure time per question was very similar to the exposure time per statement in the restudy condition. To our knowledge, our study is the first to fully control exposure time in a successive relearning experiment in a manner that is similar to that observed in the single-session testing-effect literature.³

In addition to exploring the impact of successive relearning on learning, we also investigated its impact on metacognitive accuracy with aggregate judgments of learning (aJOLs). After completion of each learning practice session, students were asked to estimate the percentage of questions on that week's topic they believed they would answer correctly on the final exam.

To address other aspects of the learning experience in the current study, we asked students to complete questionnaires at the beginning of the semester, before teaching commenced (Week 1), that measured self-reported general worry, test anxiety, academic

³ Because successive relearning necessarily is dependent on students' recall performance, the number of recall attempts before achieving mastery varied between items and between students. We limited this variation to some degree by only allowing three recall attempts per item (see Method for details). Thus, although the duration of each item exposure was fixed in the relearn condition (first pass: 20 s; second and third passes: 15 s), the number of exposures was not. Moreover, a few students failed to achieve mastery even after three recall attempts. However, as will become apparent, the constraints that we implemented were enough to control average exposure duration between the relearn and restudy conditions such that duration could not be the cause of any observed relearning (vs restudy) advantage. It is in this sense that we use the term "control".

self-efficacy, and attentional control. We also asked them to complete a visual analogue scale (VAS) at the end of each learning practice session. The VAS measured key factors that have been positively (i.e., academic efficacy, attentional control) and negatively (anxious affect) linked to achievement (reviews by Moran, 2016; Usher & Pajares, 2008).

Hypotheses and Exploratory Research Questions

Table 1 outlines 11 hypotheses and associated analyses for the study. Hypotheses 4-9 were designated central hypotheses and appear in Table 1 with an asterisk. Hypotheses 1-3 are predicted improvements to performance across sessions (within weeks) during the learning practice in the relearn condition. Specifically, across sessions we predicted an increase in recall accuracy (on the first pass through the questions; Hypothesis 1), a decrease in trials-to-criterion (Hypothesis 2), and a decrease in exposure duration (time before answering; Hypothesis 3).

Hypotheses 4-7 relate to the effect of relearning (vs. restudying) on several measures of the learning experience taken during the learning practice. Specifically, Hypothesis 4 sets out our prediction regarding aJOLs, whereas Hypotheses 5-7 pertain to the VAS ratings taken during the learning practice. Our predictions for these measures were based on the principle that, compared to restudying, relearning is the more difficult task because it involves retrieval, which would undermine the learning experience, at least initially while the task is first being performed. Therefore, judgments regarding the amount learned (Hypothesis 4), as well as feelings of anxiety (Hypothesis 5), mastery (Hypothesis 6), and attentional control (Hypothesis 7) will all be lower (higher for anxiety) in the relearn condition than in the restudy condition.

Hypotheses 8-11 specify predictions regarding recall performance on the final criterial tests. Specifically, we predicted that relearning (vs. restudying) would yield better recall performance (Hypothesis 8), that the relearning advantage (vs. restudying) would be

larger for repeated questions than transfer questions or new questions (Hypothesis 9), that recalled answers would be more accurate for repeated questions compared to transfer questions or new questions (Hypothesis 10), and that recall accuracy would deteriorate over time (Hypothesis 11).

In addition to the 11 hypotheses, we also had three additional exploratory research questions, again, with associated analyses. These questions, which were not aligned with a priori predictions, are summarized in Table 2. Specifically, Exploratory Research Questions 1-3 respectively related to (1) the relationships between aJOLs, VAS scores, learning measures during practice (recall accuracy and trials-to-criterion) and Week-1 individual difference measures, (2) whether any relearning advantage (over restudying) on the criterial tests was moderated by individual differences, and (3) the qualitative student feedback.

Method

Participants

The study consisted of two components. The first component was mandatory for students and comprised part of the coursework for the module (worth 20% of the final mark). It involved students engaging in learning practice three times per week and completing two formative tests at the end of the semester (separate from the summative final exam). Participating in the second component was voluntary but would earn research participation credit if completed (up to 2.5% of final mark). This second component consisted of (a) completing several questionnaires at the beginning of the semester and (b) providing aJOLs and completing a VAS after each learning practice session.

In the first week of the semester, all students registered on the module were sent a link that allowed them to sign up for one or both components of the study. Most of these students (93%) were under 21 years of age and most (63%) had A-level Psychology. A total of 264 students signed up for the mandatory learning practice component. Across all 30 learning

practice sessions (3 sessions per week for 10 weeks), between 199 and 247 students completed the mandatory learning practice. For the two mandatory criterial tests, 233 completed the first test and 209 completed the second test.

For the voluntary second component of the study, 256 students (97% of 264 total students) provided informed consent to complete the initial questionnaires in Week 1 and provide ratings at the end of each learning practice session for research credit. All 256 students completed at least part of the initial questionnaires. Across all 30 learning practice sessions that took place over the remainder of the semester, between 188 and 242 completed the ratings at the end of each learning session. Students who failed to complete at least 80% (24/30) of the learning practice sessions were excluded from all analyses, a constraint that we henceforth refer to as the “80% rule.” This constraint was implemented to ensure that there was sufficient opportunity for students to benefit from the learning practice. Two-hundred-and-twelve students met this criterion. The sample consisted predominately of females (191; males: 19; gender data missing: 2) and native English speakers (185; non-native English speakers: 25; speaker language data missing: 2).

Design

The design was within-participants with two experimental conditions: relearn and restudy. At the beginning of the module students were randomly allocated to one of two groups. The first group was given the opportunity to restudy the content from lectures 1, 3, 5, 7, and 9, and relearn the content from lectures 2, 4, 6, 8, and 10. For counterbalancing purposes, this task/lecture assignment was reversed for the second group. Thus, students alternated tasks each week over the 10 weeks of the study.

Learning Practice Schedule

Following the introductory week, during which students signed up for one or both components of the study, there were 10 remaining weekly lectures across the semester. We

asked students to review the content of each lecture in three learning practice sessions per week that were separated by two days (i.e., Monday: lecture; Wednesday: Practice 1; Friday: Practice 2; Sunday: Practice 3). Students reviewed the same lecture material with the same practice task (i.e., restudying or relearning) in all three learning practice sessions.

For the learning practice, students who signed up in Week 1 were sent links via email that allowed them to log in to a “Revise” website between 08:00 and 21:00 on the practice days. If the task for that week was relearning, students answered 20 fill-in-the-blank questions about that week’s lecture (e.g., _____ wrote *Principles of Psychology*). Relearning practice sessions involved mastery learning (criterion level = 1) with corrective feedback and dropout. Specifically, after indicating that they were ready to begin answering the questions by clicking an arrow, a fill-in-the-blank question was presented with a textbox below to enter a response. Regardless of what (if anything) was entered, after 15 s, the program advanced to the feedback page, displayed for 5 s. If the response was correct, the word “CORRECT!!!” was shown in green at the top of the page with the complete sentence beneath it (e.g., *James wrote Principles of Psychology*). If the response was incorrect, the word “INCORRECT!!!” was shown in red at top of the page, with the sentence containing the blank beneath it (e.g., _____ wrote *Principles of Psychology*). Beneath that, the phrase “Correct answer” followed by the solution was displayed (e.g., *Correct answer: James*). Blank responses were counted as incorrect. A page was then displayed with the instruction: “*When you are ready to proceed, please click on the arrow below.*” Thus, advancement through the questions was self-paced, while exposure duration to each question was controlled.

Once all 20 questions had been answered, either the session ended (if all questions were answered correctly) or the incorrectly answered questions were presented again. During this second pass, each question was presented for 10 s instead of 15 s (due to students being familiarized with it from the first pass), followed by 5 s of feedback as for the first pass. Only

incorrectly answered questions on the second pass were then presented for a third pass. Presentation times during the third pass were identical to the second pass. No further opportunities were provided to answer the questions correctly after the third pass. Question order was randomized for the first pass through the questions, with a fresh random order for each student. If necessary, two new random orders of the incorrectly answered questions were generated per participant for the second and third passes.

For the restudying task, the 20 questions used for the relearning task were each presented twice as statements to read. There were two presentations of each statement in the restudying task over two passes, the number of which was fixed unlike the relearning task. Each statement was presented for 20 s and 15 s during the first and second pass, respectively. Thus, the overall exposure time for a statement in the restudying task would have been exactly the same as two passes in the relearning task (including feedback). Again, a different random order of questions was used for each pass, with fresh random orders generated for each student.

Materials

Questionnaires Administered at the Beginning of the Study (Week 1)

For those students who opted in to receive research credit (see Participants section), individual differences (attentional control, academic efficacy, worry, and test anxiety) were assessed via a set of questionnaires administered once at the beginning of the semester (see Procedure section for details). Cronbach alphas across all Week-1 questionnaires were all greater than .80.

Attentional Control. We measured attentional control using the attention control scale (Derryberry & Reed, 2002). This scale consists of 20 items (nine items measure attentional focus and 11 measure attentional shifting). Individuals describe how often they experience items on a four-point scale from 1 (almost never) to 4 (always). For example, one

item is *When I need to concentrate and solve a problem, I have trouble focusing my attention.*

The possible score range is 20 – 80, with higher scores indicating poorer attentional control.

Academic Self-Efficacy. We measured academic efficacy using the self-efficacy scale (Rowbotham & Schmitz, 2013). Individuals are asked to rate how well each of 10 items describes them. Items measure the degree of confidence students have in their own ability to learn material from 1 (*not at all true*) to 4 (*exactly true*). For example, one item is *I am convinced that I am able to successfully learn all relevant subject content even if it is difficult.* The possible score range is 10 – 40 and higher scores reflect higher self-efficacy.

Worry. We measured worry using the 14-item version of the Penn State Worry Questionnaire for 6-18-year-olds (Meyer et al., 1990). Individuals are asked to indicate how well each item describes them on a four-point scale from 0 (never true) to 3 (always true), with higher values indicating more worry symptoms. For example, one item is *Many things make me worry.* The possible score range is 0 – 42 and higher scores indicating more worry.

Test Anxiety. We measured test anxiety using the test anxiety questionnaire (Nist & Diehl, 1990). Individuals are asked to rate how often they experience each of 10 items that reflect experiences of anxiety related to test-taking on a five-point scale from 1 (*never*) to 5 (*always*). For example, one item is *I feel nauseated before a test.* The possible score range is 10 – 50 with higher scores indicating more feelings of test anxiety. Scores < 20 and > 35 indicate low and high symptoms, respectively.

Learning Practice

Learning practice materials consisted of 200 questions and 200 corresponding statements (20 for each of 10 lectures). The questions were in a fill-in-the-blank format covering all aspects of the module material. Comparable statements for the restudy task were created by filling in the missing information for each question.

Criterion Test Materials

The criterial test materials consisted of two fill-in-the-blank tests, each with 30 questions. Of the 30 questions on each test, 10 were repeated, 10 were transfer questions, and 10 were new questions. The repeated questions were identical to those used for relearning practice. The transfer questions were also previously administered during relearning practice except a different part of the statement was removed. For example, one relearning practice question was “*Attitudes are composed of _____, cognitive, and behavioural components*” (correct answer: *affective*). The transfer version of this question was “*Attitudes are composed of affective, _____, and behavioural components* (correct answer: *cognitive*). Finally, the new questions queried the same general topic but were not previously included in the learning practice sessions. For example, “*False _____ is the tendency for people to believe that their own behaviour is widely shared*” (correct answer: *consensus*, for the lecture topic *Social Cognition*). Three questions on each test (one repeated, one transfer, and one new) pertained to each lecture. No question was repeated between the two criterial tests.

End-of-Session Ratings

In addition to the questionnaires administered at the beginning of the intervention, we asked students to complete a VAS (adapted from Muhl-Richardson et al., 2018) and to provide an aJOL rating at the end of each of the 30 practice sessions.

Visual Analogue Scale. This scale includes nine items that measure individuals’ current feelings. Items on the scale are associated with three core constructs: anxious affect (*anxious, worried, relaxed*), mastery (*understand, confused, mastered topic*), and attention (*focused, concentrated, disorganized*). Each item was rated on a scale from 0 to 100. Values for items pertaining to each of the three subscales were averaged to produce a single mean value for each subscale. Higher values in anxious affect indicated more anxiety, whereas higher values in attention and mastery indicated less attention and mastery, respectively. In the current sample, Cronbach alphas $\geq .69$ for each subscale across each week.

Aggregate Judgments of Learning (aJOLs). After each learning practice session, we asked students to predict their future memory performance. Specifically, participants were presented with the following statement: *I feel for Topic [weekly topic], I will answer the following percentage of questions correctly on the final exam.* Students then entered a value between 0% and 100%.

Procedure

During an introductory lecture for an *Introduction to Psychology* module, students were informed that 20% of their final mark would be based on learning practice. The practice would involve reviewing module material three times per week (30 min maximum per session) over the remaining ten weeks of the semester and completing two criterial tests at the end. They were told that they could complete the learning practice and criterial tests on any computer, tablet, or smartphone.⁴ Students were informed that for each learning practice session completed, they would receive 0.5% towards their final mark (total credit for practice = 15%: 3 practice sessions per week X 10 weeks X 0.5%). They would also receive 2.5% for completing each of the final tests (total credit for completing tests = 5%). Students were also told that the criterial tests were additional to the summative final exam and that, like the learning practice, their *performance* on the criterial tests had no bearing on the amount of module credit they received (i.e., the credit that they received was dependent on participation in the activities rather than on attainment). Thus, 20% of the final mark for the module was based on participation in this study (15% for participation in the learning practice sessions plus 5% for completing the two criterial tests).

Students were informed that they could opt in to participate in the research element of the module and earn research participation credit (as opposed to module credit). To earn this credit, students needed to complete some questionnaires and ratings in addition to

⁴ The system recorded if a mobile device was used.

participating in the learning practice sessions and the criterial tests. Specifically, four questionnaires could be completed at the beginning of the module. Also, VAS ratings and aJOLs could be completed after each learning practice session.

After the first lecture in Week 1, students received a link via email to the learning practice website (hosted by www.qualtrics.com). After logging in, students provided their gender, name, and email address. They were given the opportunity to participate in the research component by giving informed consent and completing the four questionnaires. After completing the questionnaires, students were linked to a trial relearning session to familiarize themselves with the fill-in-the-blank procedure. The trial relearning session consisted of 10 fill-in-the-blank questions about the module that were unrelated to module content (e.g., *The learning practice sessions take place on Wednesdays, Fridays and _____ throughout semester I*). These data were not analyzed.

Data collection started the following week, after the second weekly lecture. Half the students started with relearning practice, whereas the other half started with restudying practice. On each practice day, students received an email with a personalized link that allowed them to log in and engage in the learning practice when the session became available (between 8:00 and 21:00 on practice days).

After the module and learning practice was complete, students received two separate personalized emails inviting them to participate in the two criterial tests. The two tests were administered 1 day and 26 days after the final practice session (Sunday of Week 11), which meant that they were administered an average of 42 days and 68 days after first learning the material during the lecture. Consequently, we will henceforth refer to the two tests as the 42-day and 68-day tests. As with the practice sessions, students could take the tests between 8:00 and 21:00.

To evaluate the student experience associated with the learning intervention, we asked for student feedback at two different points in the semester using the same online system that administered the learning practice. We first elicited feedback toward the end of the second week of learning practice (Week 3 of the semester). The feedback prompt was “*Now that you have experienced both types of tasks, please let us know how your experience has been so far. Any feedback will be appreciated.*” The second request for feedback was distributed in the final (tenth) week of learning practice (Week 11 of the semester). The feedback prompt was “*Now that you have experienced both types of tasks for nearly the whole semester, please let us know how your experience has been. Any feedback will be appreciated.*”

Results

Overview

We divided the analysis of the results into three broad sections. In Section 1, we focus on data obtained during the learning practice sessions over the course of the semester. This section included analyses of recall accuracy, trials-to-criterion, exposure duration, aJOLs, and responses on the VAS scales (anxiety, mastery, attentional control).⁵ The primary focus was on whether there were changes in these measures across the sessions and weeks of the semester, and how they compared between the relearn and restudy conditions (except for recall accuracy and trials-to-criterion, which could not be computed in the restudy condition). Additionally, we considered (separately for the relearn and restudy conditions) associations between the Week-1 questionnaire data, the VAS scores, aJOLs, and learning measures during learning practice (i.e., recall accuracy and trials-to-criterion in the relearn condition).

⁵ Although we present the learning practice data in this section broken down by week for completeness, in most cases, it was not possible to include week as a factor in the statistical analyses because of the counterbalancing. That is, a given participant only contributed restudying data for half of the weeks, and relearning data for the other half. In these cases, the data were collapsed over week.

In Section 2, we analyzed criterial test data obtained at the end of the semester, after the learning practice was complete. The primary focus was on differences in learning between the relearn and restudy conditions as measured by the two criterial tests. We also investigated whether individual differences (i.e., the Week-1 questionnaire scores and VAS scores) were associated with any learning advantage observed in the relearn versus restudy condition. Finally, Section 3 explored the student feedback data to evaluate students' personal views of the relearn and restudy practice sessions.

The sample size varied across sessions and across weeks. As noted earlier, we applied the 80% rule to all analyses, but for the remaining participants, our strategy was to include as many of them as possible for any given analysis to maximize power. For example, on a particular week, some participants may have completed Sessions 1 and 2, but not Session 3. Rather than drop these participants altogether in analyses involving session as a factor, we estimated their Session 3 means from those participants' Session-3 data for the weeks in which they did participate. Only if no other data were available to compute a mean would participants be dropped from the analysis. However, if a participant demonstrated that they were not taking the recall task seriously by failing to answer any question correctly (or omitting all answers) on all three passes in a given session, data from that session for that participant were removed. The number of these students was very small, ranging from 0 to 5 per session ($M = 1.50$, $SD = 1.55$). Additionally, we dropped items within sessions which were recalled incorrectly on all three passes, even though the session was not dropped because other items were recalled correctly. These exclusion criteria were only applied to recall and trials-to-criterion as there was no reason to eliminate them from the analyses of the other measures (i.e., Week-1 questionnaire scores, exposure duration data, VAS scores, aJOLs judgments, and end-of-semester measures were all still valid). The precise number of

participants contributing data to any given analysis can be determined from the degrees of freedom associated with it.

We conducted sensitivity analyses for the tests of the central hypotheses (Hypotheses 4-9). Those analyses indicated that we had enough power to detect a small-to-medium size effect (or larger) for all central hypotheses (largest minimum η_p^2 detectable for any hypothesis = .044). Further details regarding the sensitivity analyses are provided in Table 1.

Section 1 - Learning Practice Data

Recall Accuracy

The first pass through the questions in the relearn condition served as an interim recall test. Consequently, it was possible to determine how recall performance on the first pass through the questions during learning practice varied over weeks and sessions within weeks (see Figure 1). To test Hypothesis 1 (Table 1), we conducted a one-way, repeated-measures Analysis of Variance (ANOVA) with session (1, 2, 3) as the independent variable. It revealed a main effect, $F(2, 422) = 2286.67, p < .001, \eta_p^2 = .916$. Tests of within-subject contrasts indicated that recall accuracy was lowest in the first session ($M = .24; SD = .14$), greatly increased in the second session ($M = .68; SD = .16, F(1, 211) = 1860.66, p < .001, \eta_p^2 = .898$), and increased further by the third session compared to the second session ($M = .82; SD = .13, F(1, 211) = 408.96, p < .001, \eta_p^2 = .660$). This pattern was observed for every week of the semester (see Figure 1). Thus, the data supported Hypothesis 1.

Trials-to-Criterion

In addition to recall, trials-to-criterion provided a second measure of learning and is shown in left panel of Figure 2 as a function of session and week. To test Hypothesis 2 (Table 1), we conducted a one-way, repeated-measures ANOVA with session as the only independent variable. It revealed a main effect, $F(2, 422) = 1863.11, p < .001, \eta_p^2 = .898$. Tests of within-subject contrasts indicated that trials-to-criterion was highest in the first

session ($M = 1.95$; $SD = .22$), greatly reduced in the second session ($M = 1.37$; $SD = .19$), $F(1, 211) = 1645.42$, $p < .001$, $\eta_p^2 = .886$, and reduced further by the third session compared to the second session ($M = 1.20$; $SD = .14$), $F(1, 211) = 286.99$, $p < .001$, $\eta_p^2 = .576$. This pattern of reduced trials-to-criterion over sessions was observed for every week of the semester. Thus, the data supported Hypothesis 2.

Exposure Duration

The number of passes through the questions in the relearn condition determined the mean duration participants were exposed to the questions in the relearn condition. These data are shown in the right panel of Figure 2 as a function of session and week.⁶ As noted earlier, exposure was set at 20 s (15 s presentation plus 5 s feedback) for the first pass through the questions, and at 15 s (10 s presentation plus 5 s feedback) for the second and third pass. (No question was presented more than three times.) To test Hypothesis 3 (Table 1), we conducted a one-way, repeated-measures ANOVA with session as the independent variable. It revealed a main effect, $F(2, 422) = 2143.00$, $p < .001$, $\eta_p^2 = .910$. Analogous to trials-to-criterion, tests of within-subjects contrasts indicated that exposure duration was highest in the first session ($M = 37.05$; $SD = 4.43$), greatly reduced in the second session ($M = 27.22$; $SD = 4.70$), $F(1, 211) = 2003.33$, $p < .001$, $\eta_p^2 = .905$, and reduced further by the third session compared to the second session ($M = 24.22$; $SD = 4.26$), $F(1, 211) = 327.11$, $p < .001$, $\eta_p^2 = .608$. This pattern of reduced exposure duration over sessions was observed for each week of the semester. Thus, the data supported Hypothesis 3.

⁶ As noted earlier, items that participants answered incorrectly on all the three passes in the relearn condition were excluded from the trials-to-criterion analysis because the criterion was not reached. However, these data were included in the analysis of exposure duration as there was no reason to exclude them (i.e., participants were still exposed to the questions and feedback even though they did not answer the questions correctly). These items each contributed the maximum value of 50 s for exposure duration. The result was that, compared to trials-to-criterion, the exposure duration curves were elevated with respect to the respective restudy control lines in Figure 2.

Our final analysis on exposure duration focused on comparing the relearn and restudy conditions. Unlike the relearn condition, the number of passes through the questions and the exposure duration on each pass was fixed in the restudy condition: specifically, statements were presented for 20 s and 15 s on the first and second pass, respectively, totaling 35 s per question. This value is shown as a dotted line in Figure 2. So that we could compare the relearn condition to the restudy condition, we summed the exposure duration in the relearn condition across passes within session for each participant and then collapsed the data over sessions and weeks. This value represented the mean total exposure time per session, which was significantly *less* ($M = 29.50$; $SD = 4.12$) than the duration in the restudy condition, $t(211) = -19.44$, $p < .001$, $d = -1.335$. The data suggest that any relearn advantage over restudy criterial test performance cannot be attributed to a confounding of practice task and exposure to the material.

Aggregate Judgments of Learning

Figure 3 shows mean aJOLs as a function of week, session, and practice task. To test Hypothesis 4 (Table 1), we conducted a 3 (session: 1, 2, 3) X 2 (practice task: relearn, restudy) repeated-measures ANOVA. It yielded a main effect of session, $F(2, 410) = 261.00$, $p < .001$, $\eta_p^2 = .560$, which was qualified by an interaction between session and practice task, $F(2, 410) = 124.49$, $p < .001$, $\eta_p^2 = .378$. Follow-up t -tests on the interaction indicated that aJOLs were higher for restudy ($M = 52.16$, $SD = 14.86$) than relearn ($M = 46.21$, $SD = 15.42$) in Session 1, $t(205) = 8.88$, $p < .001$, $d = 0.619$, there was no difference in Session 2 (restudy: $M = 55.87$, $SD = 14.50$; relearn: $M = 56.11$, $SD = 14.30$), $t(205) = -0.47$, $p = .641$, $d = -0.033$, and the difference reversed in Session 3 (restudy: $M = 58.17$, $SD = 15.17$; relearn: $M = 61.44$, $SD = 15.03$), $t(205) = -6.12$, $p < .001$, $d = -0.426$. The main effect for practice task was not significant, $F(1, 205) = 3.15$, $p = .077$, $\eta_p^2 = .015$.

Thus, Hypothesis 4 had partial support. As predicted, aJOLs in Session 1 were lower in the relearn condition compared to the restudy condition across all weeks of the semester. However, aJOLs increased in both the restudy and relearn conditions as the sessions continued, but they increased more in the relearn condition. The result was that by Session 2, the task difference in aJOLs observed in Session 1 was eliminated, and by Session 3, aJOLs were higher in the relearn condition than the restudy condition, again across all weeks of the semester.

Visual Analogue Scale

Anxious Affect. Figure 4 shows the mean anxious affect ratings (averaged over the three relevant items on the VAS) as a function of week, session, and practice task. To test Hypothesis 5 (Table 1), we conducted a 3 (session: 1, 2, 3) X 2 (practice task: relearn, restudy) repeated-measures ANOVA. It yielded a main effect of session, $F(2, 410) = 156.11$, $p < .001$, $\eta_p^2 = .432$, a main effect of practice task, $F(1, 205) = 19.57$, $p < .001$, $\eta_p^2 = .087$, both of which were qualified by an interaction between session and practice task, $F(2, 410) = 119.51$, $p < .001$, $\eta_p^2 = .368$. Follow-up t -tests indicated that anxious affect was higher for relearn ($M = 46.56$, $SD = 16.95$) than restudy ($M = 38.26$, $SD = 16.86$) in Session 1, $t(205) = 10.67$, $p < .001$, $d = 0.744$, there was no difference in Session 2 (relearn: $M = 37.31$, $SD = 16.02$; restudy: $M = 37.17$, $SD = 16.71$), $t(205) = 0.23$, $p = .819$, $d = 0.016$, and the difference reversed in Session 3 (relearn: $M = 34.18$, $SD = 16.42$; restudy: $M = 35.89$, $SD = 17.10$), $t(205) = -3.30$, $p = .001$, $d = -0.230$.

Thus, there was partial support for Hypothesis 5. As predicted, students completing the relearning task were more anxious than those who were restudying in Session 1. This difference was particularly evident in the early weeks of the semester. However, much like the aJOLs, this difference was eliminated by the second session, and by the third session, the difference had reversed.

Mastery. Figure 5 shows the mean mastery ratings (averaged over the three relevant items on the VAS) as a function of week, session, and practice task. (Higher values indicate less mastery.) To test Hypothesis 6 (Table 1), the mastery ratings were subjected to a 3 (session: 1, 2, 3) X 2 (practice task: relearn, restudy) repeated-measures ANOVA. It yielded a main effect of session, $F(2, 410) = 350.35, p < .001, \eta_p^2 = .631$, a main effect of practice task, $F(1, 205) = 6.96, p = .009, \eta_p^2 = .033$, both of which were qualified by an interaction between session and practice task, $F(2, 410) = 160.42, p < .001, \eta_p^2 = .439$. Follow-up *t*-tests indicated that students' self-rated mastery was less for relearn ($M = 55.06, SD = 12.28$) than restudy ($M = 46.68, SD = 12.97$) in Session 1, $t(205) = 10.05, p < .001, d = 0.700$, there was no difference in Session 2 (relearn: $M = 43.51, SD = 13.69$; restudy: $M = 43.89, SD = 14.50$), $t(205) = -0.62, p = .539, d = -0.043$, and the difference reversed in Session 3 (relearn: $M = 37.56, SD = 14.55$; restudy: $M = 41.11, SD = 14.64$), $t(205) = -6.00, p < .001, d = -0.418$.

These data provide partial support for Hypothesis 6. Like the anxiety ratings, students completing the relearning task rated that they had less mastery over the material in Session 1 compared to those who were restudying, a difference that was particularly noticeable in the early weeks of the semester. However, this difference was eliminated by the second session. By the third session, relearning students rated having more mastery than the restudying students, and this difference was present for every week of the semester.

Attentional Control. Figure 6 shows the mean attentional control ratings (averaged over the three relevant items on the VAS) as a function of week, session, and practice task. (Higher values indicate less attentional control.) To test Hypothesis 7 (Table 1), we conducted a 3 (session: 1, 2, 3) X 2 (practice task: relearn, restudy) repeated-measures ANOVA. It revealed a main effect of session, $F(2, 410) = 30.72, p < .001, \eta_p^2 = .130$, and a main effect of practice task, $F(1, 205) = 28.62, p < .001, \eta_p^2 = .123$. Both main effects, however, were qualified by an interaction between session and practice task, $F(2, 410) =$

67.19, $p < .001$, $\eta_p^2 = .247$. Follow-up t -tests on the interaction indicated that there was no difference between students' self-rated attentional control in the relearn ($M = 42.61$, $SD = 15.98$) versus restudy condition ($M = 41.63$, $SD = 16.43$) in Session 1, $t(205) = 1.29$, $p = .199$, $d = 0.090$. However, by Session 2, students in the relearn condition ($M = 37.92$, $SD = 15.46$) rated that they had better attentional control (restudy: $M = 42.60$, $SD = 16.67$), $t(205) = -6.79$, $p < .001$, $d = -0.473$, a difference which increased in Session 3 (relearn: $M = 36.21$, $SD = 16.06$; restudy: $M = 42.37$, $SD = 17.20$), $t(205) = -8.72$, $p < .001$, $d = -0.607$).

Thus, Hypothesis 7 was not supported by the data. Relearning and restudying students rated that they had about the same level of attentional control in Session 1. However, by Session 2, the relearning students rated that their attentional control had improved, and it improved even further by Session 3. Restudying students, on the other hand, had constant or slightly reduced attentional control over sessions.

Relationships Between aJOLs, VAS Scores, Learning Measures, and Week-1 Individual Difference Measures. To address Exploratory Research Question 1 (Table 2), we calculated mean VAS scores (separately for anxiety, mastery, and attentional control) for each participant in the relearn and restudy conditions collapsed across all learning practice sessions (i.e., each mean was based on $3 \times 5 = 15$ relearn sessions and $3 \times 5 = 15$ restudy sessions). The different VAS scores were all significantly associated with each other in both the relearn and restudy sessions, indicating that low self-reported feelings of anxiety at the completion of the practice sessions were correlated with high self-reported mastery and attentional control ($r_s > .44$ and $p_s < .001$). Table 3 shows that in both the relearn and restudy conditions, VAS scores and aJOLs were also associated with Week-1 questionnaire scores; students who reported low symptoms of test anxiety, high academic self-efficacy, and high attentional control in Week 1 also reported low VAS anxiety, high VAS attentional control and VAS mastery, as well as high aJOLs. Table 3 also shows that more worry at Week 1 was

associated with more VAS anxiety, less VAS mastery, and lower aJOLs during the practice sessions.

To further address Exploratory Research Question 1 (Table 2), we also examined the relationship between the VAS scores and measures of learning during the relearning sessions (i.e., recall accuracy and trials-to-criterion), again collapsed across all sessions. Table 4 shows that students who recalled more items and had fewer trials-to-criterion also reported reduced anxiety, more mastery and increased attentional control following the relearning practice sessions. Partial correlations between all VAS scores on the one hand and mean recall and trials-to-criterion in the relearning sessions on the other remained statistically significant after controlling for Week-1 questionnaire scores (worry, test anxiety, academic self-efficacy, and attentional control), $r_s > .24$ for trials-to-criterion and $r_s < .23$ for mean recall, $p_s < .001$ for all VAS scores.

Section 2 - End-of-Semester Data

Because there was some participant attrition by the end of the semester, one concern was that those students who wrote the criterial tests might differ from those who wrote neither test on measures taken earlier in the study (Week-1 questionnaires and learning practice measures). However, the two samples did not differ on the Week-1 individual difference measures. The samples were also comparable for most learning practice measures (recall accuracy, trials-to-criterion, exposure duration, and aJOLs). The differences that did occur related to participation and the VAS mastery judgments. Only those participants who took both tests showed a differential pattern in attendance per session, $F(2, 416) = 3.08, p = .047, \eta_p^2 = .015$, and they also showed more effect of sessions on mastery judgments on the VAS, $F(2, 404) = 3.38, p = .035, \eta_p^2 = .016$. No other differences were noted. A more complete analysis is reported as Analysis 1 in the supplemental materials.

Criterial Tests

Figure 7 shows mean proportion of repeated, transfer, and new fill-in-the-blank questions answered correctly on the two criterial tests. The left and right panels show recall performance for the day-42 and day-68 tests, respectively. To test Hypotheses 8-11 (Table 1), the data were analyzed with a 2 (practice task: relearn, restudy) X 3 (question type: repeated, transfer, new) X 2 (test delay: day-42, day-68) repeated-measures ANOVA.⁷ The analysis revealed a main effect of practice task, $F(1, 174) = 60.62, p < .001, \eta_p^2 = .258$, which supported Hypothesis 8. However, this main effect was qualified by a practice task by question type interaction, $F(2, 348) = 20.65, p < .001, \eta_p^2 = .106$. Figure 7 shows that performance was better in the relearn condition compared to the restudy condition for all question types, but the advantage was greatest for repeated questions. Follow-up t-tests indicated that the difference for repeated questions was significant, $t(174) = 9.32, p < .001, d = 0.705$ (relearn: $M = .55, SD = .23$; restudy: $M = .41, SD = .24$), while the differences for transfer (relearn: $M = .53, SD = .21$; restudy: $M = .50, SD = .22$) and new items (relearn: $M = .28, SD = .20$; restudy: $M = .25, SD = .21$) approached significance, $t(174) = 1.97, p = .051, d = 0.148$, and $t(174) = 1.94, p = .054, d = 0.147$, respectively. This pattern of data supported Hypothesis 9.

The analysis also yielded main effects of question type, $F(2, 348) = 326.39, p < .001, \eta_p^2 = .652$, test delay, $F(1, 174) = 23.11, p < .001, \eta_p^2 = .117$. The main effect of question type supported Hypothesis 10. The main effect of test delay was opposite to Hypothesis 11, as explained below. The main effects of question type and test delay were qualified by a

⁷ Due to a technical error, for the first few hours of the day-68 test, 72 students were incorrectly administered day-42 test rather than the day-68 test. After the error was detected and corrected, these students were invited to login again and take the day-68 test. Fifty-two students chose to do so. To account for any variance attributable to this error (e.g., different participant characteristics, performance advantages of taking the day-42 test a second time shortly before taking the day-68 test, and so on), we separated these students' data from the rest of the sample by adding a "test administration" (correct, incorrect) between-subjects factor to the ANOVA. Importantly, the test administration factor did not interact with any other factors and the pattern of significance for the factors other than test administration was the same for both analyses. Consequently, here we report the simpler analysis not including test administration. However, the analysis that included it, as well as the results comparing the two analyses, are reported as Analysis 2 in the supplemental materials.

significant two-way interaction, $F(2, 348) = 3.71, p = .026, \eta_p^2 = .021$. This interaction was due to the second versus first test advantage being greatest for new questions (day-42 test: $M = .22, SD = .19$; day-68 test: $M = .32, SD = .25$), medium for repeated questions (day-42 test: $M = .46, SD = .23$, day-68 test: $M = .51, SD = .26$), and least for transfer questions (day-42 test: $M = .49, SD = .21$; day-68 test: $M = .53, SD = .24$). Follow-up t -tests indicated that the advantage was significant for all three question types, $t(174) = -5.13, p < .001, d = -0.388$, $t(174) = -2.84, p = .005, d = -0.215$, and $t(174) = -2.14, p = .034, d = -0.162$, for new, repeated, and transfer questions, respectively. No other main effect or interaction was significant, although the three-way interaction was marginally significant, $F(2, 348) = 2.57, p = .078, \eta_p^2 = .015$, which occurred because the relearn (vs. restudy) advantage for repeated items was somewhat larger for the day-42 test than the day-68 test.

Relationships Between VAS Scores, and Criterial Tests, and Week-1 Individual Difference Measures

To address Exploratory Research Question 2 (Table 2), we computed a difference score between the number of repeated test items answered correctly from the relearn practice sessions and those answered correctly from restudy practice sessions. Increased scores indicate more benefit of relearning versus restudying material, a benefit we refer to as the *relearn advantage*. We then determined whether either set of individual differences measures (VAS scores and Week-1 questionnaires) were associated with the relearn advantage. The correlations between the relearn advantage scores and the VAS scores are reported in Table 5. None of these correlations was significant. Further analyses also confirmed that the relearn advantage was not associated with Week-1 questionnaires scores either ($r_s < .10, p_s > .10$). Together, these results suggest that the relearn advantage was not associated with individual differences reported either at the starting point of learning (i.e., beginning of the semester) or during the learning practice.

Section 3 - Student Feedback Data

Our final analysis addressed Exploratory Research Question 3 (Table 2). Student feedback about the learning practice obtained at two points of the semester (Weeks 2 and 10) sent a consistent message: compared to the restudying task, the relearning task was strongly preferred as students reported it to be more effective, beneficial, engaging, active, helpful, and enjoyable. Students complained that they found it difficult to concentrate during the restudying task and that it was not helpful for learning. Regarding the relearning task, several complaints pertained to not having enough time to type in their answers before the website progressed to the feedback page, and having answers marked as incorrect because of minor errors (e.g., misspelled words). However, overall, the prevalent theme in the feedback was that the relearning task was seen to be an effective study technique and it was strongly preferred to restudying.

Discussion

Benefits to Recall Accuracy

In the current research, we implemented successive relearning in an introductory psychology class, and it yielded multiple learning benefits. First, compared to restudying, successive relearning produced clear benefits to memory retention of the course material. It is highly unlikely that the learning benefits that we observed on memory retention in the relearn condition occurred because participants were exposed to or thought about the material for longer than in the restudy condition. Our study was the first to our knowledge that has attempted to control exposure duration between the restudy and relearn conditions in a manner like that seen in the single-session testing-effect literature. Rawson et al. (2013, Experiment 1) included a restudy control condition in their research, but exposure in that condition was self-regulated and, although the number of restudy trials was fixed, there was the opportunity for a further restudy trial if participants so chose. Thus, there was potential

between-subject and between-item variability in both the number of times an item was restudied and in the duration of the restudy exposure on a given trial.

Like Rawson et al. (2013), we determined restudy exposure duration by estimating how many times on average it would take participants in the relearn condition to reach criterion. However, we eliminated the sources of variability that were present in Rawson et al.'s research by strictly controlling the number of restudy trials per item (fixed at two) and the exposure duration on each trial (20 s on the first pass and 15 s on the second). Given the feedback that students provided, we believe it was a good decision to not allow students to self-regulate their restudy times; the feedback indicated that restudying was by far less popular than relearning, so if given a choice, students would likely have chosen to restudy items for considerably less time than relearning them.

Our a priori estimate of two trials-to-criterion in the relearn condition led to the decision to expose each statement in the restudy condition for 35 s (20 s and 15 s for the first and second pass, respectively); in fact, average trials-to-criterion in the relearn condition was somewhat lower than two such that, on average, each question was exposed for less than 30 s. This low exposure duration in the relearn condition remained true even though we retained trials for which students failed to produce a correct answer after three passes. These trials, which were eliminated from the trials-to-criterion dataset because criterion was not reached, represent the maximum possible exposure duration (50 s) and their inclusion elevated the relearn exposure duration means across all weeks relative to the trials-to-criterion means (cf. left and right panels in Figure 2). Thus, our estimate of exposure duration in the relearn condition was liberal. The fact that overall relearn exposure duration of each item was less in the relearn versus restudy condition, coupled with the fact that some of the question (the part left blank) was not exposed until the final 5 s on each trial, strongly suggests that at least

some of the learning benefits of successive relearning are attributable to factors other than mere exposure.

That said, some aspects of our data suggest that the size of successive relearning effects reported in the prior literature may have been considerably overestimated by failing to control for exposure. For example, consider the recall accuracy patterns shown in Figure 7. Specifically, recall performance for the new items was considerably lower than it was for the other item types. In previous studies without transfer items, the efficacy of relearning would be determined by comparing repeated-relearned items to new (or “business-as-usual”) items rather than to repeated-restudied items (as in our study). The former comparison is substantially larger than the latter, suggesting that repeated restudying, in and of itself, benefited learning when compared to new items, which is exactly the type of exposure effect that we controlled in our study.

Figure 7 shows that the test administered, on average, 68 days after learning the material during the lecture (right panel) had better recall than the test administered after 42 days (left panel). On the surface, this result may be surprising because usually forgetting causes recall to deteriorate as retention intervals get longer. However, the likely reason for better recall is that the 68-day test was closer to the summative final exam which was scheduled for ten days later. Consequently, a significant portion of students had likely started preparing for the final exam by day 68, which would account for the high recall accuracy. More importantly for present purposes, the fact that recall in the relearn condition was higher than in the restudy condition regardless of whether students had been learning the course material on their own time suggests that the learning benefits of successive relearning are highly robust.⁸

⁸ We further explored the effect of retention interval on recall accuracy by examining the relearn advantage on the criterial tests for material learned in the first four weeks of the semester versus the last four weeks. Semester time did not moderate the relearn advantage, $F(1, 174) = 0.33, p = .568, \eta_p^2 = .002$. The full results of the analysis are reported as Analysis 3 in the supplemental materials.

Benefits to Metacognition

Importantly, the learning benefits that successive relearning produced were not just limited to memory retention of repeated items. Successive relearning also had a beneficial effect on metacognitive monitoring as measured by aJOLs. During the initial learning practice session, participants rated that they would remember the information contained in the questions and answers worse than if the statements were presented intact for restudying. By the second session, this difference was eliminated (restudy = relearn), and by the third session, the difference reversed (relearn > restudy).

Had we stopped at one session per week, the higher aJOLs given to restudied statements versus answered questions (with feedback) would be consistent with previous literature showing that participants do not monitor the memorial benefits of testing very well (e.g., Karpicke, 2009; Kornell & Son, 2009; Potts & Shanks, 2014; Roediger & Karpicke, 2006). For example, Potts and Shanks (2014) had participants learn foreign vocabulary pairs in two experimental conditions: participants either studied the pairs intact (read condition) or they first guessed the English word translation and then received corrective feedback (generate condition). Because the pairs were new to participants, the guesses in the latter condition were nearly always wrong. Across a series of experiments, they found that JOLs assigned in the read condition were consistently higher than those assigned in the generate condition. Presumably, the act of generating multiple errors led participants to believe that they were not learning very much. However, later memory performance showed the opposite pattern. That is, *errorful generation* followed by corrective feedback was better for learning than studying the intact pairs, despite participants believing the opposite.

We did not test memory at the end of the semester for information that only underwent one learning practice session, so we were unable to explore the possibility of a recall/JOL dissociation analogous to that observed by Potts and Shanks (2014). However, it is

worth noting that recall accuracy in the first session of our study was quite poor in the relearn condition, just as it was with Potts and Shanks. This poor performance may have been the source of the low aJOLs. However, as recall accuracy improved in the second and third sessions, aJOLs increased accordingly, such that by the third session, aJOLs were in line with recall (i.e., both showed a relearn advantage). This finding is reminiscent of other research that has found that participants' metacognition can be "trained" to accurately reflect the potency of retrieval practice if they are given the opportunity to experience that potency and there is external guidance (e.g., see Tullis et al., 2013; see also Barenberg & Dutke, 2019; Miller & Geraci, 2014 for other metacognitive benefits of testing). This result is promising in that it suggests that successive relearning provides the necessary support to correct faulty metacognitions about retrieval practice.

Benefits to Anxious Affect, Mastery, and Attentional Control

The benefits of successive relearning were not limited to just memory retention and more accurate metacognition. Compared to restudying, successive relearning also reduced students' anxious affect, increased their subjective sense of mastery over the to-be-learned material, and improved their self-reported attentional control. These benefits were not realized immediately; overall in Session 1, relearning was inferior to restudying on these measures. However, after the third successive relearning session each week, the benefits of relearning (vs. restudying) were clearly observed, as with aJOLs.

We attribute these delayed anxiety, mastery, and attentional control benefits to improvements in recall performance over sessions: recall in Session-3 was over three times better than Session-1 recall (Figure 1). Poor Session-1 recall of answers to questions about a lecture held only two days earlier likely gave these students, who were only in their first year at university, a sense that they were not learning very much. However, as recall improved

with more relearning sessions, so did students' ratings of anxiety, mastery and attentional control

High anxiety and poor mastery ratings in Session 1 of the relearn condition were particularly evident, at least descriptively, in the first six weeks of the learning practice. However, as the semester progressed, these extremely negative Session-1 ratings eased off, presumably because students realized as the learning practice routine was established that the first recall task would be hard, but it would get easier in the later sessions.

Together, these data patterns demonstrate the benefit of *repeating* retrieval practice over multiple sessions. Had we administered just one learning practice session per week, we would have reached a very different conclusion about the potential value of providing practice questions for students versus giving them the opportunity to restudy key lecture points. Compared to restudying, relearning just once meant that students predicted that they would not do well on later tests, found it to be anxiety provoking, did not get a sense of mastery, and reported little attentional control. It was only by continuing to retrieve answers to the same set of questions over multiple sessions, the hallmark of successive relearning (along with feedback and mastery), that the benefits were realized.

Individual Differences

We conducted several correlational analyses to assess the degree to which individual difference measures taken in Week 1 and during the learning practice (VAS scale) were associated with other variables. Overall, most of these individual difference measures were associated with each other. Additionally, they were associated with our measures of metacognition (aJOLs) and learning (recall and trials-to-criterion; see Tables 3 & 4). The association between individual differences such as anxiety, mastery, attentional control, and self-efficacy on the one hand and learning on the other is broadly consistent with prior research (see reviews by Moran, 2016; Usher & Pajares, 2008). However, our research

extends previous work by demonstrating that repeated successive relearning sessions reduced feelings of anxiety, whilst increasing self-reported feelings of mastery and attentional control. Moreover, our results showed that the relearning learning advantage over restudying was not associated with any individual differences measure (see Table 5). The value of successive relearning might be undermined if, say, only non-anxious students, or those with high academic self-efficacy benefited from successive relearning, whereas other students did not. However, our results suggest that there are benefits from successive relearning for all students. This finding is consistent with recent research which has shown that individual differences have no effect on other beneficial effects of low-stake testing, such as the facilitation of new learning (i.e., forward testing effect; Yang et al., 2020). Therefore, the adoption of successive relearning methods in teaching may help to address challenges associated with engagement and achievement for students who experience negative affect and lowered attentional control (e.g., Owens et al., 2014; Steinmayr et al., 2016).

Limitations

One limitation of our research is that the memorial benefits of successive relearning (over restudy) were limited to repeated questions. Although there were marginally significant successive relearning advantages to transfer and new questions (compared to restudying), they had small effect sizes. This small effect size was understandable for the new questions because their only relationship to the practiced questions was that they pertained to the same general topic (e.g., clinical psychology). However, it was quite surprising that the transfer questions did not show a reliable benefit; these questions were the same as those answered during practice except that participants were required to fill in a different part of the question. The concern here is that successive relearning merely helped students to memorize specific answers to specific questions rather than fostering “deep” learning or a general understanding of the topic at hand. One potential reason for this specificity might be that our questions

queried detailed information and may not engender processes that would be conducive to transfer. For example, learning the names of psychology's pioneers does little to assist students in learning other course material, even if it is on the same general topic (historical foundations). Future research should investigate ways for successive relearning to enhance the transfer of learning to new material. For example, different types of questions might be used and/or students might be required to provide reasons (or to choose reasons from a list of options) for why the answers they are providing are correct. Nonetheless, the multiple additional benefits to affect and metacognition still make successive relearning an excellent learning tool even if the memorial benefits were limited in this study.

A second limitation, which applies to all field research, is that we had little control over what participants were doing during the learning practice or during the criterial tests. Because course credit was dependent on participation rather than attainment, a handful of students logged in to "complete" the relearning practice but failed to produce correct answers even after three attempts. Occasionally, this occurred for all questions within a session. Similarly, some restudying students may have disengaged from the task, not reading the statements but merely advancing the trials after the allotted exposure time was complete. Importantly, though, if this did occur, it did not appear to occur too often; the systematic data patterns across multiple dependent variables suggest that most students took the tasks seriously, no doubt to enhance their own learning in the module.

Conclusions

Our research has shown that successive relearning is a powerful learning tool that is easy to implement in real classrooms and that it confers multiple learning benefits. Similar to prior research (e.g., Bahrick, 1979; Bahrick et al., 1993; Bahrick & Hall, 2005; Rawson et al., 2013; Rawson & Dunlosky, 2011, 2012, 2013; Vaughn et al., 2016), we found that successive relearning produced good recall of to-be-learned material over both 42-day and 68-day (on

average) retention intervals. However, our research is the first to our knowledge to demonstrate this excellent performance in comparison to a restudy control condition with rigorously controlled exposure to the to-be-learned material. Our research was also unique in that we explored the effect of successive relearning on a host of educationally relevant variables in addition to memory retention. Our findings indicated that successive relearning was beneficial to metacognition, anxiety, mastery, and attentional control. Overall, our research suggests that students should be encouraged to incorporate successive relearning into their learning, either by making it a coursework component, as we did here, or assigning it as homework. Critically, however, it is important to ensure that whatever form successive relearning takes, it must involve multiple opportunities for students to practice retrieving the same material. Otherwise, some of the learning benefits may be limited or even negative.

References

- Agarwal, P. K., D'Antonio, L., Roediger, H. L., McDermott, K. B., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory and Cognition*, 3(3), 131–139. <https://doi.org/10.1016/j.jarmac.2014.07.002>
- Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, 21, 627–635. <https://doi.org/10.1080/00140137808931764>
- Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, 108, 296–308. <http://dx.doi.org/10.1037/0096-3445.108.3.296>
- Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science*, 4(5), 316–321. <https://doi.org/10.1111/j.1467-9280.1993.tb00571.x>
- Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language*, 52(4), 566–577. <https://doi.org/10.1016/j.jml.2005.01.012>
- Barenberg, J., & Dutke, S. (2019). Testing and metacognition: Retrieval practise effects on metacognitive monitoring in learning from text. *Memory*, 27(3), 269–279. <https://doi.org/10.1080/09658211.2018.1506481>
- Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, 61(3), 228–247. <https://doi.org/10.1016/j.cogpsych.2010.05.004>
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354–380. <https://doi.org/10.1037/0033-2909.132.3.354>

- Derryberry, D., & Reed, M.A. (2002). Anxiety-related attentional biases and their regulation by attentional control. *Journal of Abnormal Psychology*, 111, 225-236.
<https://doi.org/10.1037/0021-843X.111.2.225>
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Emeny, W. G., Hartwig, M. K., & Rohrer, D. (2021). Spaced mathematics practice improves test scores and reduces overconfidence. *Applied Cognitive Psychology*, acp.3814.
<https://doi.org/10.1002/acp.3814>
- Healy, A. F., Jones, M., Lalchandani, L. A., & Tack, L. A. (2017). Timing of quizzes during learning: Effects on motivation and retention. *Journal of Experimental Psychology: Applied*, 23(2), 128–137. <https://doi.org/10.1037/xap0000123>
- Janes, J. L., Dunlosky, J., Rawson, K. A., & Jasnow, A. (2020). Successive relearning improves performance on a high-stakes exam in a difficult biopsychology course. *Applied Cognitive Psychology*, 34(5), 1118–1132. <https://doi.org/10.1002/acp.3699>
- 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., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1250–1257.
<https://doi.org/10.1037/a0023436>
- Kornell, N., Klein, P. J., & Rawson, K. A. (2015). Retrieval attempts enhance learning, but retrieval success (versus failure) does not matter. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 41(1), 283–294.

<https://doi.org/10.1037/a0037850>

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory*,

17(5), 493–501. <https://doi.org/10.1080/09658210902832915>

Logan, J. M., Castel, A. D., Haber, S., & Viehman, E. J. (2012). Metacognition and the

spacing effect: The role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal. *Metacognition and Learning*, 7(3), 175–195. <https://doi.org/10.1007/s11409-012-9090-3>

Lyle, K. B., Bego, C. R., Hopkins, R. F., Hieb, J. L., & Ralston, P. A. S. (2020). How the amount and spacing of retrieval practice affect the short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, 32(1), 277–295.

<https://doi.org/10.1007/s10648-019-09489-x>

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State worry questionnaire. *Behaviour Research and Therapy*, 28(6), 487–495. [https://doi.org/10.1016/0005-7967\(90\)90135-6](https://doi.org/10.1016/0005-7967(90)90135-6)

Miller, T. M., & Geraci, L. (2014). Improving metacognitive accuracy: How failing to retrieve practice items reduces overconfidence. *Consciousness and Cognition*, 29, 131–140. <https://doi.org/10.1016/j.concog.2014.08.008>

Moran, T.P. (2016). Anxiety and working memory capacity: A meta-analysis and narrative review. *Psychological Bulletin*, 142, 831–864. <https://doi.org/10.1037/bul0000051>

Muhl-Richardson, A., Godwin, H. J., Garner, M., Hadwin J. A., & Liversedge, S. P. & Donnelly, N. (2018). Individual differences in search and monitoring for color targets in dynamic visual displays. *Journal of Experimental Psychology: Applied*, 24, 564–577. <https://doi.org/10.1037/xap0000155>.

- Nist, P., & Diehl, M. (1990). Test anxiety questionnaire anxiety questionnaire. Retrieved September 18, 2020, from <https://www.yumpu.com/en/document/view/51745094/test-anxiety-questionnaire>.
- Owens, M., Stevenson, J., Hadwin, J. A., & Norgate, R. (2014). When does anxiety help or hinder cognitive test performance? The role of working memory capacity. *British Journal of Psychology*, 105(1), 92–101. <https://doi.org/10.1111/bjop.12009>
- Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, 143(2), 644–667. <https://doi.org/10.1037/a0033194>
- Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60(4), 437–447. <https://doi.org/10.1016/j.jml.2009.01.004>
- Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General*, 140(3), 283–302. <https://doi.org/10.1037/a0023956>
- Rawson, K. A., & Dunlosky, J. (2012). When is practice testing most effective for improving the durability and efficiency of student learning? *Educational Psychology Review*, 24(3), 419–435. <https://doi.org/10.1007/s10648-012-9203-1>
- Rawson, K. A., & Dunlosky, J. (2013). Relearning attenuates the benefits and costs of spacing. *Journal of Experimental Psychology: General*, 142(4), 1113–1129. <https://doi.org/10.1037/a0030498>
- Rawson, K. A., Dunlosky, J., & Sciartelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychology Review*, 25(4), 523–548. <https://doi.org/10.1007/s10648-013-9240-4>

- Rawson, K. A., Vaughn, K. E., Walsh, M., & Dunlosky, J. (2018). Investigating and explaining the effects of successive relearning on long-term retention. *Journal of Experimental Psychology: Applied*, 24(1), 57–71. <https://doi.org/10.1037/xap0000146>
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science*, 17(3), 249–255.
- Rowbotham, M., & Schmitz, G.S. (2013) Development and validation of a student self-efficacy scale. *Journal of Nursing & Care*, 2: 126. <https://doi.org/10.4172/2167-1168.1000126>
- 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>
- Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(4), 907–912. <https://doi.org/10.1037//0278-7393.27.4.907>
- Steinmayr, R., Crede, J., McElvany, N., & Wirthwein, L. (2016). Subjective well-being, test anxiety, academic achievement: Testing for reciprocal effects. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01994>
- Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, 110(16), 6313–6317. <https://doi.org/10.1073/pnas.1221764110>
- Tullis, J. G., Finley, J. R., & Benjamin, A. S. (2013). Metacognition of the testing effect: Guiding learners to predict the benefits of retrieval. *Memory & Cognition*, 41(3), 429–442. <https://doi.org/10.3758/s13421-012-0274-5>

- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, 48(4), 751-796.
<https://doi.org/10.3102/0034654308321456>
- Vaughn, K. E., Dunlosky, J., & Rawson, K. A. (2016). Effects of successive relearning on recall: Does relearning override the effects of initial learning criterion? *Memory & Cognition*, 44(6), 897–909. <https://doi.org/10.3758/s13421-016-0606-y>
- Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: A review of the forward testing effect. *Npj Science of Learning*, 3(1).
<https://doi.org/10.1038/s41539-018-0024-y>
- Yang, C., Sun, B., Potts, R., Yu, R., Luo, L., & Shanks, D. R. (2020). Do working memory capacity and test anxiety modulate the beneficial effects of testing on new learning? *Journal of Experimental Psychology: Applied*. <https://doi.org/10.1037/xap0000278>
- Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society*, 15(1), 41–44. <https://doi.org/10.3758/BF03329756>

Table 1

Hypotheses and Associated Analyses for the Study. Sensitivity is Also Shown for Each of the Central Hypotheses ().*

Hypothesis	Analysis
Hypothesis 1: Recall accuracy on the first pass through the questions during relearning practice will increase over learning sessions within week.	One-way, repeated-measures ANOVA on recall accuracy (first iteration) with session (1, 2, 3) as the independent variable.
Hypothesis 2: Trials-to-criterion will decrease over learning sessions within week.	One-way, repeated-measures ANOVA on trials-to-criterion with session (1, 2, 3) as the independent variable.
Hypothesis 3: Exposure duration during learning practice will decrease over learning sessions within week.	One-way, repeated-measures ANOVA on exposure duration with session (1, 2, 3) as the independent variable.
*Hypothesis 4: aJOLs will be lower in the relearn condition than in the restudy condition.	3 (session: 1, 2, 3) X 2 (practice task: relearn, restudy) repeated-measures ANOVA on aJOLs. Sensitivity (η_p^2) = .037.
*Hypothesis 5: Anxious affect will be higher in the relearn condition than in the restudy condition.	3 (session: 1, 2, 3) X 2 (practice task: relearn, restudy) repeated-measures ANOVA on anxious affect scores (VAS). Sensitivity = (η_p^2) = .037.
*Hypothesis 6: Mastery will be lower in the relearn condition than in the restudy condition.	3 (session: 1, 2, 3) X 2 (practice task: relearn, restudy) repeated-measures ANOVA on mastery scores (VAS). Sensitivity (η_p^2) = .037.

*Hypothesis 7: Attentional control will be lower in the relearn condition than in the restudy condition.	3 (session: 1, 2, 3) X 2 (practice task: relearn, restudy) repeated-measures ANOVA on attentional control scores (VAS). Sensitivity (η_p^2) = .037.
*Hypothesis 8: Recall performance on the criterial tests will be better in the relearn condition than the restudy condition.	2 (practice task: relearn, restudy) X 3 (question type: repeated, transfer, new) X 2 (test delay: day-42, day-68) repeated-measures ANOVA on recall accuracy. Sensitivity (η_p^2) = .044.
*Hypothesis 9: The relearn advantage (over restudying) for recall performance on the criterial tests will be larger for repeated questions compared to transfer questions or new questions.	2 (practice task: relearn, restudy) X 3 (question type: repeated, transfer, new) X 2 (test delay: day-42, day-68) repeated-measures ANOVA on recall accuracy. Sensitivity (η_p^2) = .028.
Hypothesis 10: Recall performance on the criterial tests will be better for repeated questions compared to transfer questions or new questions.	2 (practice task: relearn, restudy) X 3 (question type: repeated, transfer, new) X 2 (test delay: day-42, day-68) repeated-measures ANOVA on recall accuracy.
Hypothesis 11: Recall performance on the criterial tests will be better for the Day-42 test than the Day-68 test.	2 (practice task: relearn, restudy) X 3 (question type: repeated, transfer, new) X 2 (test delay: day-42, day-68) repeated-measures ANOVA on recall accuracy.

Note. Central hypotheses are indicated with an asterisk (*). Sensitivity reported with the analyses that tested each of the central hypotheses specifies the minimum effect size (η_p^2) detectable with the sample size associated with the analysis, assuming $\alpha = .05$ and power = .80.

Table 2*Exploratory Research Questions and Associated Analyses*

Exploratory Research Questions	Analysis
Exploratory Research Question 1: What are the relationships between aJOLs, VAS scores, learning measures during practice (recall accuracy and trials-to-criterion) and Week-1 individual difference measures?	Correlational analyses.
Exploratory Research Question 2: Is the relearn advantage (vs. restudy) for recall performance on the criterial tests moderated by individual differences.	Correlational analysis.
Exploratory Research Question 3: How does qualitative student feedback compare between restudying versus relearning?	No formal analysis.

Table 3

Mean (SD and Actual Range) Week-1 Questionnaire Scores (Worry, Test Anxiety, Attention, and Self-Efficacy) and Associations with the Mean Self-Reported Visual Analogue Scale (VAS) Scores and Mean Aggregate Judgments of Learning (aJOLs)

Week-1 measure	VAS scores and aJOLs										
	Descriptive statistics			Relearn practice sessions				Restudy practice sessions			
	Mean	SD	Range	Anxiety	Attention	Mastery	aJOL	Anxiety	Attention	Mastery	aJOL
Worry	24.68	8.43	5-40	.34***	.10	.18**	-.15*	.34***	.15*	.18**	-.14*
Test anxiety	31.23	7.55	10-50	.39***	.29**	.42***	-.42***	.40***	.28**	.36***	-.37***
Attention	49.48	7.03	32-70	.22**	.34***	.37***	-.33***	.23***	.38***	.37***	-.37***
Self-efficacy	32.06	3.77	23-40	-.21**	-.31***	-.25**	.30***	-.22**	-.33***	-.28***	.35***

Note. High VAS scores indicate judgments of high anxiety and predictions of high future test scores (aJOLs), but low attention and mastery. High Week-1 questionnaire scores indicate poor attentional control, but high self-reported symptoms of worry, test anxiety, and self-efficacy. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4

Mean (SD and Actual Range) Recall Accuracy and Trials-to-Criterion (TtC) During Relearning Practice and Associations with Mean Self-Reported Visual Analogue Scale (VAS) Scores and Mean Aggregate Judgments of Learning (aJOLs)

Learning measure	Descriptive statistics			Mean VAS scores and aJOLs			
	Mean	SD	Range	Anxiety	Attention	Mastery	aJOL
Recall	.55	.14	.02 - .87	-.31**	-.26*	-.40***	.21*
TtC	1.54	.18	1.14 - 2.40	.27*	.24**	.35**	-.16**

Note. High scores indicate high self-reported anxiety and high aJOLs, but low attention and mastery. Mean recall is the mean proportion of questions correctly recalled across all relearn practice sessions and trials-to-criterion reflects the mean number of recall attempts to reach criterion across all relearn practice sessions. * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 5

Mean (SD and Actual Range) Relearn Advantage for Repeated Items on the 42-Day and 68-Day Tests and Associations with the Mean Self-Reported Visual Analogue Scale (VAS) Scores.

Test delay	Descriptive statistics			VAS scores		
	Mean	SD	Range	Anxiety	Attention	Mastery
42-Day Test	.16	.31	-.10 - .80	-.07	.03	-.01
68-Day Test	.10	.26	-.60 - .80	-.08	.04	-.08

Note. High VAS scores indicate judgments of high anxiety and predictions of high future test scores (aJOLs), but low attention and mastery. The relearn advantage represents the difference between the number of correctly recalled repeated versus restudied items; higher scores indicate a greater relearn advantage.

Figure 1

Mean Relearn Recall Accuracy on the First Pass Through the Questions During Learning Practice for Each Week and Session. The Textual Ms and SDs Shown Are Collapsed Across Weeks. Error Bars Represent ± 1 Standard Error from the Mean.

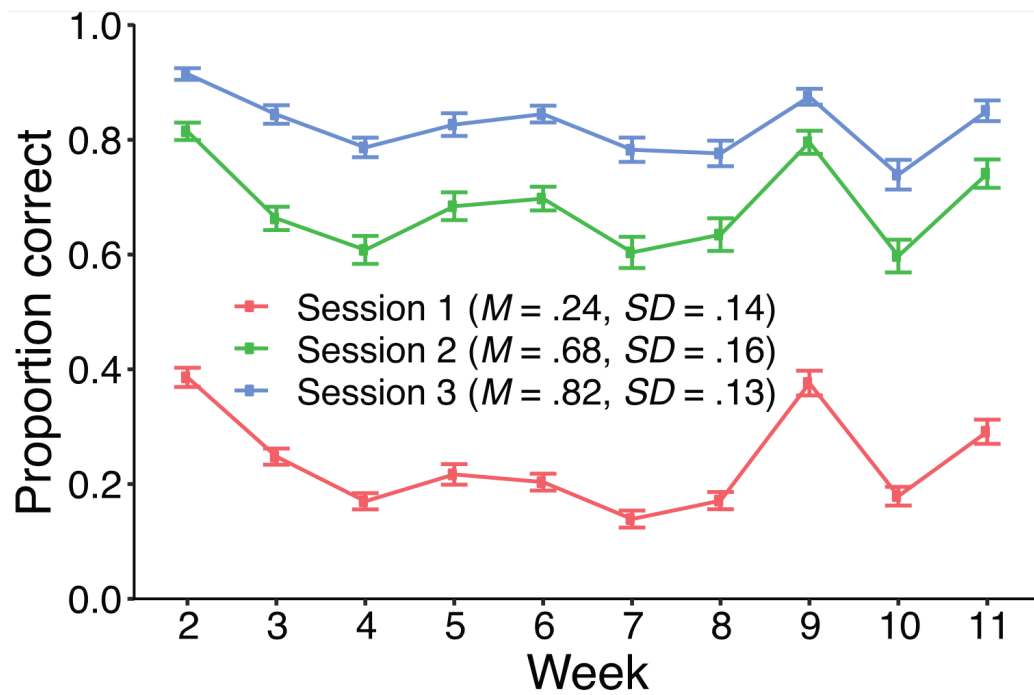


Figure 2

Mean Trials-to-Criterion and Exposure Duration (s) in Learning Practice for the Relearn Condition by Session and Week. Dashed Lines Represent the Values for the Restudy Condition. The Textual Ms and SDs Shown Are Collapsed Across Weeks. Error Bars Represent ± 1 Standard Error from the Mean.

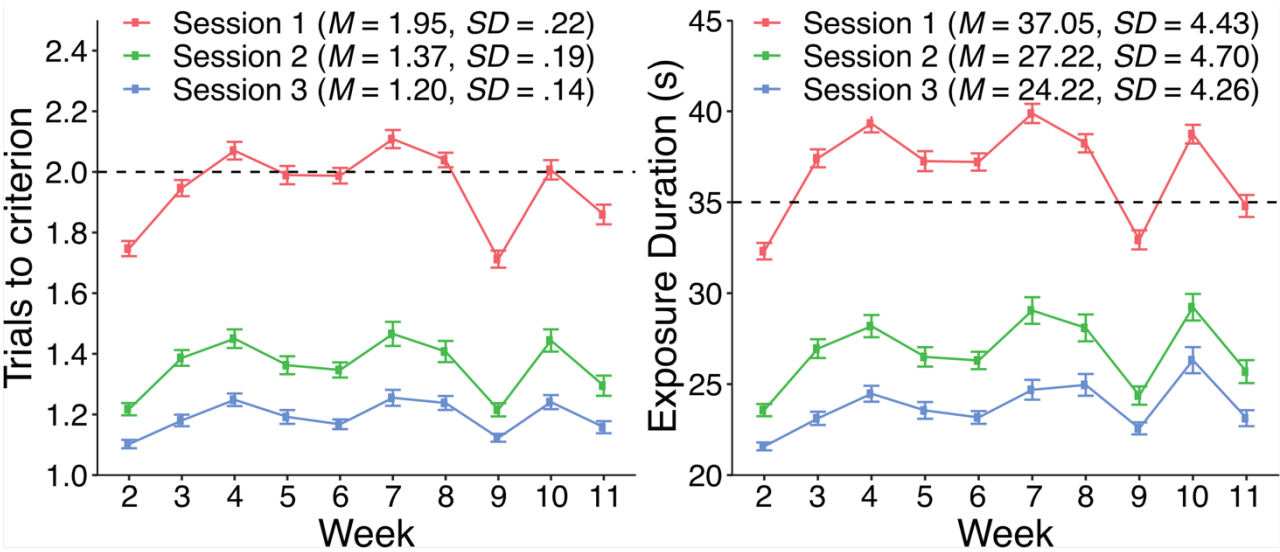


Figure 3

Mean Aggregate Judgments of Learning (0-100) During Learning Practice for Each Week, Practice Type and Session. The Textual Ms and SDs Shown Are Collapsed Across Weeks. Error Bars Represent ± 1 Standard Error from the Mean.

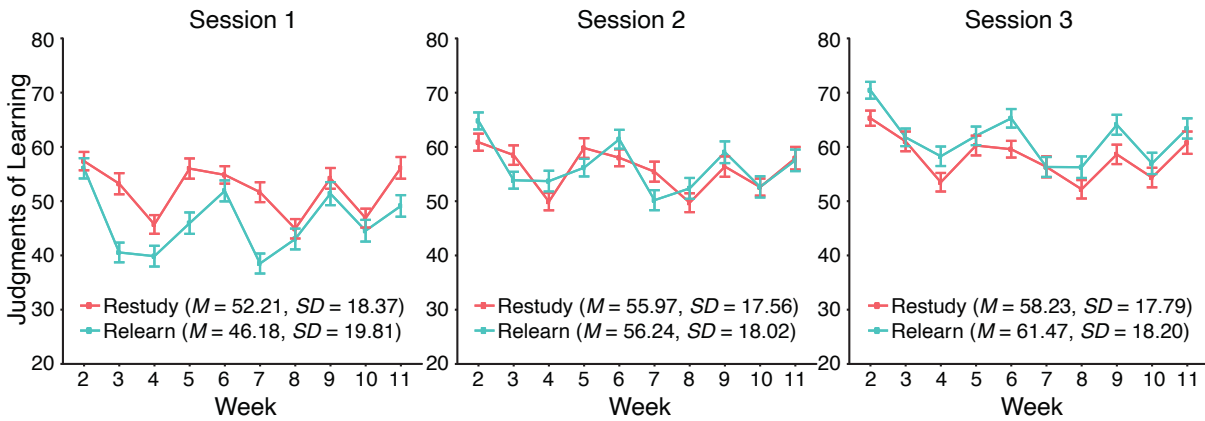


Figure 4

Mean Anxious Affect from The Visual Analogue Scale for Each Week, Practice Type, and Session. The Textual Ms and SDs Shown Are Collapsed Across Weeks. Error Bars Represent ± 1 Standard Error from the Mean.

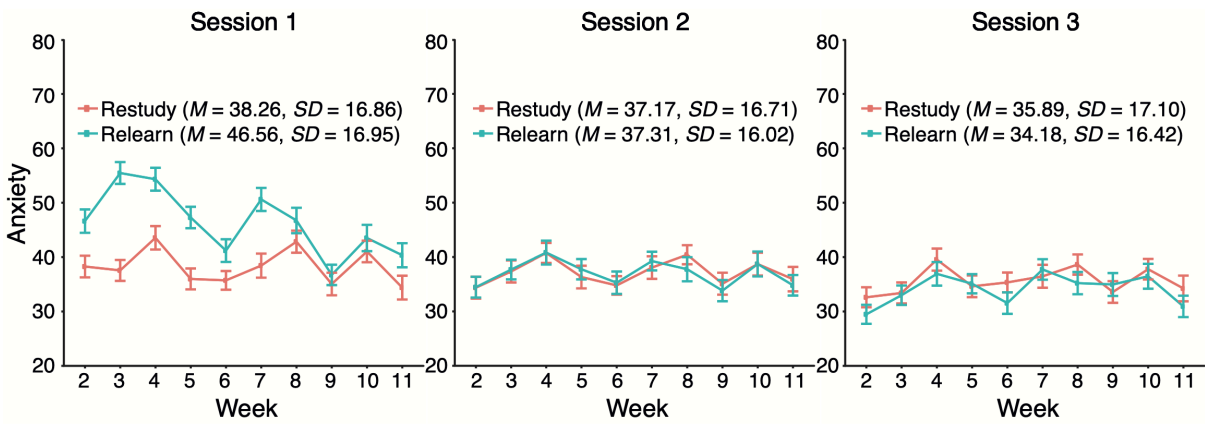


Figure 5

Mean Mastery from The Visual Analogue Scale for Each Week, Practice Type, and Session. Higher Scores Indicate Less Mastery. The Textual Ms and SDs Shown Are Collapsed Across Weeks. Error Bars Represent ± 1 Standard Error from the Mean.

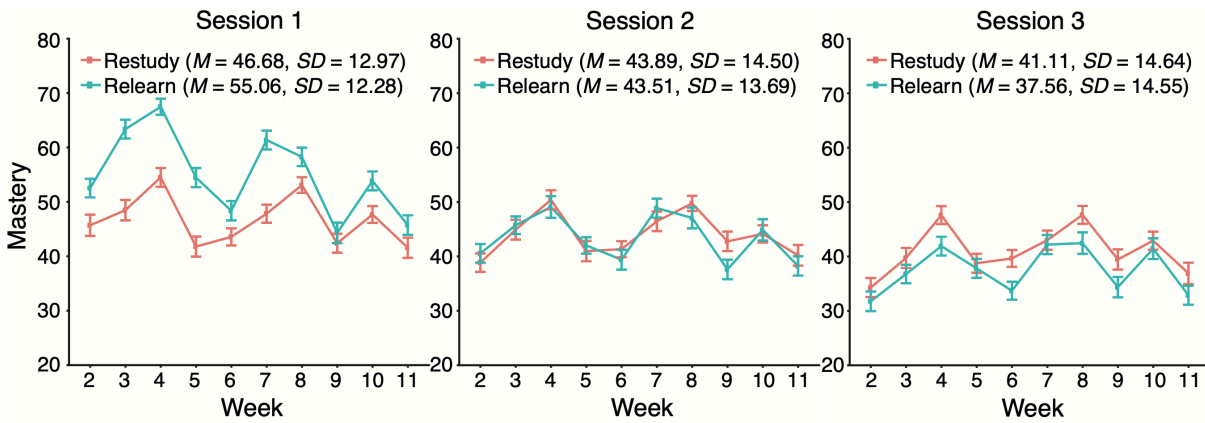


Figure 6

Mean Attentional Control from the Visual Analogue Scale for Each Week, Practice Type, and Session.

Higher Scores Indicate Less Mastery. The Textual Ms and SDs Shown Are Collapsed Across Weeks.

Error Bars Represent ± 1 Standard Error from the Mean.

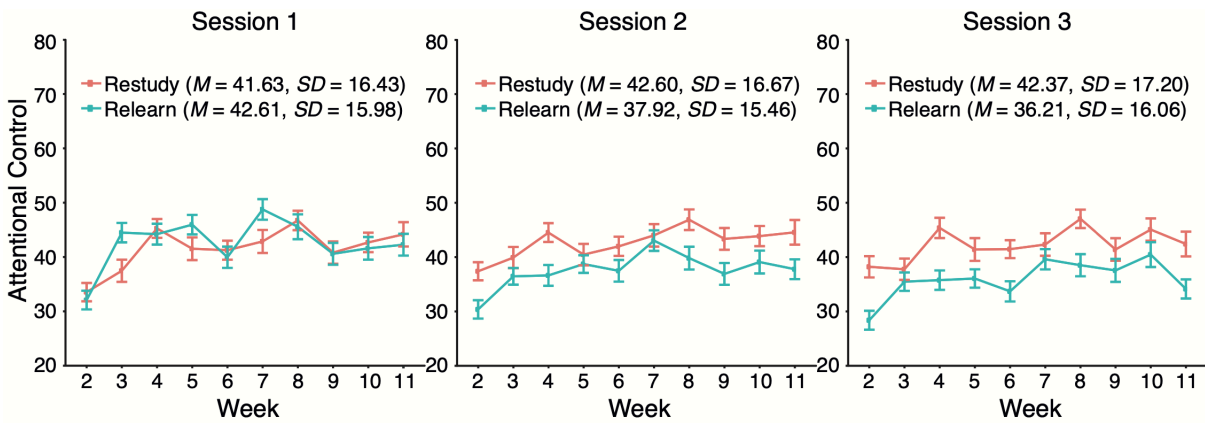


Figure 7

Mean Percentage of Answers Correctly Recalled on the Two Criterial Tests for Each Practice Type and Question Type. The Left and Right Panels Show Recall Accuracy for the 42-Day and 68-Day Tests, Respectively. Error Bars Represent ± 1 Standard Error from the Mean.

