SUCCESSIVE RELEARNING IMPROVES PERFORMANCE ON A HIGH-STAKES EXAM IN A DIFFICULT BIOPSYCHOLOGY COURSE

A thesis submitted

To Kent State University in partial

Fulfillment of the requirements for the

Degree of Master of Arts

by

Jessica Lynn Janes

May, 2019

Thesis Written by

Jessica Lynn Janes

B.S., Colorado State University, 2015

M.A., Kent State University, 2019

Approved by

Dr. John Dunlosky_____________________, Advisor

Dr. Maria Zaragoza____________________, Chair, Department of Psychological Sciences

Dr. James Blank______________________, Dean, College of Arts and Sciences

LIST OF TABLES

LIST OF FIGURES

Acknowledgments

First and foremost, I would like to thank John Dunlosky for his mentorship, constructive feedback, and help in designing and conducting the current experiments. I would also like to thank Katherine A. Rawson for providing her input and expertise throughout the course of this study. In addition, I would like to thank Aaron Jasnow for his collaborative efforts in conducting this study. Finally, I would like to thank Maria Zaragoza, Manfred Van Dulmen, and Mary Beth Spitznagel for serving on my thesis committee and providing me with helpful feedback during my defense.

Successive Relearning Improves Performance on a High-Stakes Exam in a Difficult Biopsychology Course

In sports, music, and many other domains, coaches and instructors often invoke the phrase "practice makes perfect" – and for a good reason. A basketball player attempting to improve his or her free-throw average may need to spend some time each day shooting from the line until he or she has made a certain number of shots. A musician struggling to remember a certain verse of a song may need to rehearse those lyrics on numerous occasions until he or she can sing them with ease. In such cases, acquisition of a new skill involves practicing until some level of mastery has been reached on multiple occasions. Despite the widespread use of this approach to learning in numerous real-world contexts, learning in educational contexts is often approached differently (Rawson, Vaughn, Walsh, & Dunlosky, 2018). That is, many students report preferring strategies that do not encourage mastery or yield durable learning, such as rereading or cramming the evening before an exam (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). These observations raise the question of whether students should instead be practicing their learning of course materials in a manner similar to that of practicing skills in other contexts, which involves practicing to a criterion of success on multiple occasions – a technique that is referred to as *successive relearning*.

In the current study, we evaluated the degree to which implementing successive relearning in a difficult biopsychology course would promote student learning as measured by performance on a high-stakes exam. In the following section, we first describe the successive relearning method and briefly review prior work that has examined its effectiveness. Next, we discuss a major limitation of prior research – namely, whether the benefits of successive relearning generalize to authentic classroom settings. Finally, we overview how we investigated successive relearning in the present classroom experiments.

Successive relearning refers to *multiple successful retrievals* that are *distributed across sessions* (e.g., Bahrick, 1979). In particular, successive relearning involves alternating between retrieval and study attempts until a certain level of retrieval success has been met (e.g., a criterion of one successful retrieval), and then relearning that material (by again alternating between retrieval and study) to criterion in subsequent sessions. This method is analogous to a student practicing a stack of flashcards until each of the to-be-learned items can be correctly retrieved from memory, and then returning to relearn the content in that stack on multiple occasions. As may be evident from this description, successive relearning combines two highly effective learning techniques – retrieval practice and spaced practice (for recent reviews, see McDaniel & Little, in press; Wiseheart, Kupper-Tetzel, Weston, Kim, Kapler, & Foot, in press). As described further below, research on successive relearning in authentic classroom contexts is limited, but laboratory studies demonstrate how powerful and long-lasting its effects could be for promoting student learning.

To illustrate the potential power of successive relearning, consider one noteworthy study by Bahrick, Bahrick, Bahrick and Bahrick (1993). Over the course of 9 years, the four authors investigated their retention of 300 foreign language translations across several conditions (each of which involved a different combination of the number of relearning sessions and the spacing between those sessions). Each relearning session began with a cued-recall test over a given set of items. Each learner presented themselves with each foreign word on one side of an index card and attempted to recall its English translation on the other. When an item had been correctly retrieved, the card was dropped from practice for that session. When an item had been incorrectly retrieved, they reviewed the item on the back of the card and set it aside for another test trial later in that session. Any given relearning session was complete once each of the translations had been correctly retrieved. Final retention tests were administered 1, 2, 3, or 5 years following the last relearning session. Considering the length of these delays, levels of final retention were impressive. For instance, in one condition in which relearning sessions were spaced at a 56-day interval, more than 75% of the items were recalled after a 1-year delay, and more than 60% of them were still retained after a 5-year delay (collapsed across the number of relearning sessions). Other investigations of successive relearning have yielded similarly impressive levels of final memory performance (Bahrick, 1979; Bahrick & Hall, 2005; Dunlosky, Rawson, & Sciartelli, 2013; Pyc & Rawson, 2011; Rawson & Dunlosky, 2011, 2013; Rawson et al., 2018; Vaughn, Dunlosky, & Rawson, 2016).

Despite such promising outcomes, the majority of prior work has involved learning conditions that are not representative of those in which students typically learn. Rawson et al. (2013) noted that some of the earlier investigations of successive relearning involved time scales that are implausible to instantiate in academic contexts. In Bahrick et al. (1993), for example, participants practiced material as many as 26 times spanning the length of four years, an untenable schedule for students enrolled in a typical course spanning the length of only four months. Moreover, the majority of successive relearning studies have been conducted in controlled laboratory settings using simple materials. In fact, out of the 9 successive relearning studies conducted to date, 6 of them used foreign language vocabulary or some form of paired associate (Bahrick, 1979; Bahrick & Hall, 2005; Bahrick et al., 1993; Pyc & Rawson, 2011; Rawson et al., 2018; Vaughn et al., 2016), whereas only 3 used materials more representative of the types of content students might be expected to learn in their courses (key-term definitions; Rawson & Dunlosky, 2011, 2013; Rawson et al., 2013). Given these limitations, an open question is whether the benefits of successive relearning would occur when implemented in an authentic classroom setting, where the stakes tend to be higher and the students may be more motivated to learn the material. Participants in the prior laboratory studies were not able to return to the materials following successive relearning, whereas students often study the night before an exam (Taraban, Maki, & Rynearson, 1999). This raises the interesting possibility that certain regulatory decisions made by students (e.g., cramming) may override the benefits of successive relearning on a high-stakes exam.

This possibility has been evaluated in only a single investigation examining the influence of successive relearning on performance in a classroom setting (Rawson et al., 2013). Students enrolled in Introductory Psychology successively relearned key concepts from their course via a virtual flashcard program. Practice sessions were aligned with the course schedule, such that students began practicing content during the week in which it was introduced in lecture. Half of the concepts were practiced with successive relearning, whereas the other half were not and served as a business-as-usual control. Performance on the course exam was more than a letter grade higher for questions that tapped practiced versus control concepts (84% vs. 72%). Moreover, students retained 64% of the practiced concepts when given a cued-recall test 24 days later. Thus, successive relearning was an effective means of enhancing exam performance and promoting long-term retention of the course material.

Outcomes from Rawson et al. (2013) are promising, but as noted above, that was the only investigation to examine the efficacy of successive relearning in an authentic educational context. A goal of education research, however, is to support informed decisions about which techniques to adopt and to base recommendations on empirical evidence collected in representative contexts. According to standards created by the *What Works Clearinghouse* (WWC; a division of the US Department of Education's Institute of Education Sciences that serves as a source of scientific evidence for what works in education), it may be difficult to estimate the effectiveness of an intervention with evidence from only a single investigation, given the potential for sampling error and other characteristics to have influenced the results (WWC, 2013). To better examine the utility of a technique, a larger number of investigations, range of participants, and range of settings is recommended. Accordingly, a clear need exists to examine the efficacy – and potential boundary conditions – of successive relearning across a wide range of contexts to establish the size of such effects and ultimately determine its educational value.

The current work critically extends prior research by investigating the effects of a successive relearning intervention in a biopsychology course, as measured by performance on a high-stakes exam. We chose this course because it requires students to learn a large amount of material, and we wanted to investigate the potential for successive relearning to promote achievement in a course with which many students struggle. Doing so will provide important evidence into the potential utility and generalizability of the technique. In two experiments, students earned homework credit for engaging in successive relearning to learn key concepts from their course that later appeared on their first exam. Students were provided with a virtual flashcard program that contained several to-be-learned topics, each of which served as a different "stack" of flashcards containing four to six items. Depending on the nature of the topic, students engaged in three different types of flashcard practice (see Figures 1-3). Some stacks involved retrieving key-term definitions, others involved labeling key parts of an image (e.g., structures in the brain), and others involved retrieving key information at each step in a given process (e.g., the sequence of events in an action potential). Students were provided with a homework schedule that indicated when each stack should be completed. The schedule for practicing the topics of the stacks was designed to align with the course schedule such that the stacks served as a supplement to the content being presented in the textbook and class lectures. Critically, only some of the topics that were drawn from the course were covered in the stacks, which allowed us to compare students' exam performance for those topics relative to those learned on their own (i.e., business as usual). Moreover, we divided the topic stacks available on the program into two sets, and students were only assigned to successively relearn one set for their homework. However, students had access to the other set so that they could use successive relearning to practice all concepts if they wished. Finally, to further explore the influence of successive relearning, we compared exam grades from students in the current investigation to those from students who took the course in a prior semester without the successive relearning program.

Experiment 1

Methods

Participants. Fifty undergraduate students enrolled in an upper-division biopsychology course participated as part of their regular class activities (i.e., for homework credit). Two students did not take the first unit exam, resulting in a final sample of 48 students. Demographic information was not requested from students.

Materials and design. Materials included 88 concepts drawn from two units that the instructor covered in class prior to the first exam (neuroanatomy and neurophysiology). The concepts fit into five general topics, three of which came from the neuroanatomy unit and two of which came from the neurophysiology unit. In the order of when those topics were introduced in class, they included (1) the organization of the nervous system and major brain structures, (2) functions of major brain structures, (3) cells within the nervous system, (4) key-terms relating to neurophysiology, and (5) neurotransmission. As illustrated in Table 1, we separated the concepts within each of those topics as evenly as possible among three sets: Set A, Set B, and Baseline. For both Sets A and B, each of the five topics included a different virtual stack of flashcards (for topic names, see Table 1). The two stacks for a given topic were similar across the two sets, such that they consisted of a similar number of to-be-learned concepts and always involved the same type of practice. That is, for each set, two stacks required retrieving key-term definitions, two required labeling parts of an image, and one involved retrieving key information at each step in a given process (described further in the procedure). Both sets were placed on flashdrives and one flashdrive was given to each student. Across flashdrives, we counterbalanced the assignment of the two sets to two different practice conditions. For half of the flashdrives, Set A was assigned to a successive relearning condition (and placed in a folder labeled "Required") whereas Set B was assigned to an optional practice condition (placed in a folder labeled "Optional"), and vice versa for the other half of flashdrives. The baseline set of concepts was not included on the flashdrive. Thus, our design included three conditions: successive relearning (SR), optional practice, and baseline control.

Each flashdrive also included an electronic copy of a homework schedule that detailed three different days on which each of the required stacks should be completed (see Table 2). This schedule also included a cover page that provided students with information regarding the purpose of the assignments (i.e., about successive relearning) along with detailed instructions on how to use the flashcard program.

The exam contained 33 multiple-choice questions covering the two units previously discussed. Fifteen of those questions targeted baseline concepts whereas the other 18 targeted concepts included on the flashcard program, with a similar number of questions targeting Sets A and B. Example questions can be found in Appendix A. In addition to the multiple-choice questions, the exam included 17 short-answer questions. However, the vast majority of those questions targeted Set A concepts, and thus we do not consider them further.

Procedure. During the second class period of the semester, the first and second author visited the class to distribute the flashdrives. The programs were distributed in an alternating fashion based on students' seating arrangements, such that for every two students in a row of seats, one received a flashdrive in which Set A was assigned to the SR condition, and the other received a flashdrive in which Set B was assigned to the SR condition. Students were also provided with a hard copy of the homework schedule included on their flashdrives. Next, the second author explained how to use the program, the specifics of the assignments, and the importance of using the program with fidelity (i.e., to gain the benefits of successive relearning). Students were also encouraged to practice the optional set and to continue studying course content as they normally would. They were informed that the successive-relearning programs were merely a supplement to their learning in the course, and that they only contained a subset of the material that would be evaluated on the first course exam. Students were told that they could use the flashdrive wherever they had access to a computer or laptop (the library, at home, etc.). If they did not have access to a computer or laptop, they could request one from us (although no student did so).

An overview of the homework schedule is provided in Table 2. Students were assigned a total of six homework assignments, with consecutive assignments spaced apart by an average of three days. For the first assignment, students were instructed to practice the three stacks corresponding to the topics in the neuroanatomy unit (refer to Table 1 for the particular stacks). The stacks did not need to be completed in any given order. As described in detail below, the assignment was considered complete once each of the to-be-learned concepts (within each of the stacks assigned) had been correctly retrieved. Students received credit for the assignment once they emailed the first author three files (one for each stack) that appeared on the flashdrive following practice. The second assignment required practice of the same three stacks (again until each of the to-be-learned concepts had been correctly retrieved) for the second time. The third assignment required practice of those same three stacks for the third time. For the fourth assignment, students practiced for the first time the two stacks corresponding to the two topics in the neurophysiology unit. The fifth and sixth assignments required practice of those two stacks again for the second and third time, respectively. Thus, students were required to complete each of the five stacks a total of three times and thus were expected to send 15 files in total to receive full credit for the homework (students received partial credit if they did not send all 15 files). Students took the exam the day after the last homework assignment was due.

When any given stack was opened on the flashdrive, students were instructed to type in a username of their choosing (to be used consistently across stacks). Next, they were shown an instruction screen reminding them of the given task. Given that students engaged in three different types of flashcard practice depending on the topic of the stack, we describe each one of them in turn. For stacks involving key-term definitions, students were told that they would engage in three activities (retrieving, scoring retrieved responses, and restudying). Students were told that they would practice recalling definitions until they correctly retrieved each one and that the computer would be scoring the accuracy of their responses (although recall accuracy was actually based on students' self-score judgments). Students were then presented with the concepts one at a time for a self-paced *retrieval-monitoring-feedback* (RMF) trial. Each trial began with retrieval practice, in which a term appeared at the top of the screen and students attempted to type its definition into the text box below (see the left panel of Figure 1). Once students indicated they were done retrieving their answer (by clicking a button labeled "done with this answer"), they were prompted to score the accuracy of their retrieval response for that term. As shown in the middle panel of Figure 1, the term was presented at the top of the screen, and the correct definition was broken down into its main ideas, each of which included "yes" and "no" buttons alongside it. Students were instructed to compare each of the ideas to their response (displayed below the main ideas) to indicate whether their response contained each idea (for details on idea-unit judgments, see Dunlosky, Hartwig, Rawson, & Lipko, 2011). To enhance the accuracy of the judgments, on trials in which a student indicated that they had recalled all of the ideas but their response had fewer than half the number of characters as in the correct definition, the program instructed them to revisit their judgment, as it was incorrect. Next, a feedback screen presented the term and its full definition for restudy (see the right panel of Figure 1).

On any given RMF trial, if students had not scored their response as entirely correct (by clicking "yes" to each of the main ideas), that term was placed at the end of the stack for another trial later in that session. If students had judged their response as entirely correct, that term was dropped from practice for that session. Once each of the key-term definitions had been judged as correct, the program informed students that they were done for that set of material for the day and that a data file should have appeared on their flashdrive. Note that students could exit the program at any time and a data file would still appear, which could be viewed to examine their progress for that session.

For stacks involving an image, students were instructed that they would be shown an image and that they would need to label (from memory) the key parts of that image. They were also told that they would be given feedback and would need to score their response as correct or incorrect. Following the instructions, students were presented with an image and attempted to label various parts of that image one at a time (see Figure 2). For each trial, a line pointed to a particular area of the image along with the prompt "Name of structure?", and students typed their response into the text box below. Once students indicated that they were done responding, the correct answer appeared below their response, and they indicated whether their response was right or wrong by clicking "correct" or "incorrect." (Note: Because the names of the structures were relatively short, students did not compare their responses with idea-unit feedback as they did during key-term definitions practice.) Students could study the correct answer as long as they wanted before making their judgment. Once the judgment was made, students were prompted to label a different part of the image. As with key-term definitions practice, if students scored their response as correct, that part of the image was dropped from practice. Students were done with practice for that session once all parts of an image were correctly retrieved.

For stacks involving a process, students were given the same instructions as with image practice, except that they were told they would be presented with a diagram illustrating a given process and that they would be filling in key information at each step in that process (rather than labeling structures). As illustrated in Figure 3, on each trial, a line pointed to a given step in the process (e.g., the first event in an action potential), and below that line was a statement that illustrated that step with some part of it missing (e.g., the cell is at rest, so voltage-gated ____?____ are closed). Students typed their response in the text box below. As with image practice, once students had typed their response, the correct answer appeared below it and students indicated whether it was correct or incorrect. Again, students were done with practice for that stack once each of the concepts (here, the key information at each step) had been correctly retrieved.

Results and Discussion

Four of the students did not complete a single homework assignment and thus presumably did not use the flashdrive to successively relearn. Those students also never returned their flashdrives following the exam, and so it was impossible to discern whether they had been assigned to practice set A or B. Nonetheless, we did not want to exclude those students from our analyses, and so we randomly assigned each of them to one of the two sets for the analyses. (Note that this intent-to-treat analysis may have led to conservative estimates of any successive relearning advantage.) Given that our primary interest was to compare exam performance for questions targeting SR concepts (i.e., those required for homework) relative to questions targeting optional or baseline concepts, we separated performance on the exam questions into three categories (SR, optional, and baseline) and collapsed our data across concept sets.

Exam Performance. Mean performance on the exam is presented in Figure 4. A repeated-measures analysis of variance (ANOVA) revealed a significant difference in exam performance among the three types of questions, $F(2, 94) = 11.76$, $p < .001$, $\eta_p^2 = .20$. Planned comparisons revealed that performance was higher for questions tapping SR than baseline concepts, $t(47) = 5.02$, $p < .001$, $d = .54$, and was higher for questions tapping SR than optional concepts, $t(47) = 3.41$, $p < .001$, $d = .66$. Performance did not differ between questions tapping optional and baseline concepts, $t(47) = .35$, $p = .725$, $d = .05$.

Cohort Comparison. To further investigate the potential benefit of successive relearning, we compared students' performance on the exam to that of students from the prior Spring semester (Figure 5). These students had the same professor, experienced the same lectures (to the best of the professor's ability), and took the same exam. The only difference between them was the implementation of the successive relearning intervention. If successive relearning enhanced exam performance for the experimental class relative to the comparison cohort, then we would expect the cohorts to have similar performance on baseline control items and for a difference to emerge on SR items. (Note that because we were interested in the specific influence of successive relearning on performance, we did not include performance on optional items in this analysis.) Consistent with this expectation, a 2 (cohort: experimental vs. control) x 2 (question type: successive relearning vs. baseline) mixed ANOVA revealed a significant interaction, $F(1, 105) = 5.79$, $p = .019$, $\eta_p^2 = .05$. Performance on the baseline concepts did not differ between cohorts, $t(106) = 1.92$, $p = .06$, $d = .37$, but performance on SR concepts was greater for the experimental than comparison cohort, $t(105) = 3.64$, $p < .001$, $d = .71$. A main effect of question type also emerged, $F(1, 105) = 28.22$, $p < .001$, $\eta_p^2 = .21$, and so did a main effect of cohort, $F(1, 105) = 9.47$, $p = .003$, $\eta_p^2 = .08$.

In sum, successive relearning enhanced exam performance relative to optional practice and business-as-usual activities, boosting students' scores by a letter grade (approximately 10%). A similar improvement (approximately 12%) was observed when students that engaged in successive relearning were compared to those from a prior semester that did not. Taken together, these findings demonstrate that successive relearning can be an effective means of promoting student learning and achievement on authentic course assessments.

Experiment 2

Outcomes from Experiment 1 were promising, and given the importance of direct replications (Rodgers & Shrout, 2018; Simons, 2014), the primary goal of Experiment 2 was to replicate the findings from Experiment 1. However, one limitation of Experiment 1 is that we only evaluated the effectiveness of successive relearning on a single exam. Thus, another goal of Experiment 2 was to extend the findings from Experiment 1 by having students engage in successive relearning prior to the first and second exams. Prior to both exams, students successively relearned course content via virtual flashcard practice and again earned homework credit. Students were given a similar number of stacks to practice for the second exam as with the first exam, with similar spacing between the assignments. Unfortunately (and unbeknownst to the professor), a copy of the second exam was obtained by some students prior to the exam date, rendering data from Exam 2 uninterpretable. Due to this issue, in Experiment 2, we report a direct replication of Experiment 1 and do not discuss details of the second exam further.

Methods

Participants. Twenty-six undergraduate students enrolled in an upper-division Biopsychology course participated in the study as part of their regular class activities (i.e., for homework credit). Four students did not take the first unit exam, resulting in a final sample of 22 students. Demographic information was not requested from the students.

Materials, design, and procedure. The same materials and design were used from Experiment 1. The procedure was also the same, with two minor exceptions: we distributed the flashdrives to students during the first (rather than second) class period of the semester, and students took the exam three days (rather than 1 day) following the last homework assignment. These changes were made only to ensure the assignments were spaced apart by an average of 3 days (as in Experiment 1). The due dates of the assignments are listed in Table 2.

Results and Discussion

As in Experiment 1, we separated exam performance into three categories (SR, optional, or baseline) and collapsed our data across concept set for the following analyses.

Exam performance. Mean performance on the exam is presented in Figure 6. A repeated-measures ANOVA revealed a significant difference in exam performance among the three types of questions $F(2, 42) = 8.46$, $p < .001$, $\eta_p^2 = .29$. Planned comparisons revealed that performance was greater for questions tapping SR than baseline concepts, $t(21) = 4.32$, $p < .001$, $d = 1.10$. Performance was also greater for questions tapping optional than baseline concepts, $t(21) = 2.60$, $p = .017$, $d = .55$. Performance did not differ between questions tapping SR and optional concepts, $t(21) = 1.56$, $p = .134$, $d = .41$.

Cohort comparison. To be consistent with Experiment 1, we compared students' exam performance to that of students who took the course in a prior semester (see Figure 5). We again expected the cohorts to have similar performance on baseline control items but for a difference to emerge on SR items. Consistent with this expectation, a 2 (cohort) x 2 (question type) mixed ANOVA revealed a significant interaction, $F(1, 79) = 12.21$, $p = .001$, $\eta_p^2 = .13$, such that performance on the baseline concepts did not differ between cohorts, $t(80) = .291$, $p = .77$, $d = .07$, whereas performance on the SR concepts was higher for the experimental than comparison cohort, $t(79) = 2.70$, $p = .01$, $d = .67$. A main effect of question type also emerged, $F(1, 79) = 31.30$, $p < .001$, $\eta_p^2 = .28$, but the main effect of cohort did not, $F(1, 79) = 1.89$, $p = .17$, $\eta_p^2 = .02$.

Consistent with outcomes from Experiment 1, successive relearning produced nontrivial gains in learning: it enhanced exam performance by a letter grade and a half (approximately 16%) when compared to business-as-usual activities and by a letter grade (approximately 11%) when compared to students from a prior semester that did not successively relearn. Inconsistent with Experiment 1, optional practice also enhanced exam performance relative to business-as-usual activities – a finding we return to in the General Discussion.

To what extent did students engage in successive relearning?

We should note that in the current experiments, implementing successive relearning required a nontrivial amount of time and effort among students. For instance, students had to complete homework assignments on a relatively frequent basis (every 3 days or so), requiring them to plan and budget their time accordingly. Furthermore, students needed to be at least somewhat motivated to attempt the relatively large number of required assignments (they needed to submit 15 files total). As such, we were interested in the extent to which students were able to follow homework instructions and complete their assignments in a manner consistent with the schedule. As evident in Table 3, students in both experiments completed the majority of the assignments required for homework, earning an average grade of 91%. Furthermore, the majority of those assignments were completed on time (on or before the due date), suggesting that students followed the homework schedule fairly well.

A related question concerns the degree to which students were able to use the flashcard programs with fidelity – that is, to what extent did students actually engage in successive relearning? Recall that successive relearning is defined as *multiple successful retrievals* that are *distributed across sessions*. Two outcomes of interest are thus (1) the degree to which students reached criterion (defined as 1 successful retrieval per concept) during a given practice session and (2) the degree to which students spaced their practice across sessions. Concerning the former, reaching criterion depended on students' monitoring judgments – a concept would be dropped from practice only once the student had judged it as correct (despite students being told that the program would track their progress). Thus, if students were not accurate in grading their responses, practice may have been terminated prematurely, potentially limiting the effectiveness of successive relearning. To explore this possibility, we computed the mean number of times a definition was correctly retrieved during practice. As evident in Table 3, across sessions students reached an average criterion of .87. This suggests that students may have been slightly overconfident when grading their responses, consistent with prior work showing that idea-unit judgments provide good but not perfect monitoring accuracy (e.g., Dunlosky et al., 2011). Nevertheless, a criterion of .87 is not far from the target criterion of 1.

Concerning spacing, although students were encouraged to master each of the stacks on three separate occasions, they had the flexibility to schedule their practice as they wished. This means that even if students submitted their assignments every few days (to comply with the homework schedule), it does not necessarily mean that they spaced apart their practice with any given stack. For instance, students may have decided to practice a stack again after just having mastered it rather than return to master it again a few days later. Hence, we computed the average number of days between practice sessions with any single stack (based off time stamps from students' files rather than the date the assignments were submitted). Consistent with the spacing of the due dates on the homework schedule, students spaced apart their practice with a given stack by a few days on average (see Table 3). Taken together, these findings suggest that students complied with instructions and implemented successive relearning with fidelity.

General Discussion

The goal of the current study was to implement successive relearning into a biopsychology course and evaluate its effectiveness at enhancing student performance on a high-stakes exam. Outcomes from both experiments were impressive: compared to business-as-usual activities, successive relearning enhanced students' exam performance by an average of 13%. Moreover, a comparable boost in performance was observed when students that engaged in successive relearning were compared to those from a prior semester that did not. Consistent with outcomes from Rawson et al. (2013), these findings demonstrate that the benefits of successive relearning extend to authentic educational contexts and suggest that successive relearning is an effective way to enhance student learning.

One unexpected finding was that optional practice (relative to business-as-usual activities) enhanced learning in Experiment 2 but not in Experiment 1. This difference may be explained by differing levels of baseline performance between experiments (performance on optional items was similar across experiments). That is, any learning gains that resulted from students' optional practice in Experiment 1 may not have emerged due to relatively high baseline levels among those students. One reason why baseline performance could be slightly elevated in Experiment 1 (relative to Experiment 2) is that the professor may have conveyed some of the information in a slightly different or more effective manner during that particular semester, despite his best efforts to keep the lectures consistent across semesters. Another possibility is that students in Experiment 1 were more motivated to learn the material or used more effective study strategies than did students in Experiment 2. In either case, the difference in baseline levels between experiments was nonsignificant, $p = .061$, $d = .42$. More important, baseline performance in both experiments did not differ from that of the comparison cohort.

To what extent did students engage in optional practice prior to the first exam? As a reminder, students were told that the optional set of materials was as important as the set required for homework – both sets contained the same number of stacks and would be equally represented on their exam. Nonetheless, the majority of students (61% across both experiments) did not engage in any optional practice whatsoever, which is especially surprising considering students were equipped with a tool that would allow them to do so fairly easily. Moreover, among the students who did engage in optional practice, the total number of practice sessions engaged in was relatively low and varied considerably across students ($M = 3.78$; $SD = 6.03$). Importantly, this number was considerably lower than that targeted by the successive relearning condition, which was a total of 15 sessions (5 stacks 3 times each). Obviously, the extent to which students might engage in optional practice will depend on several factors, such as the amount of time students have to dedicate to such practice as well as their willingness to do so. However, as discussed above, in the current experiments students needed to be willing to spend a nontrivial amount of time and effort on the homework assignments. That students were able to implement successive relearning when the assignments were required but engaged in substantially less practice under optional conditions highlights an important point with respect to practicality: in the absence of an external incentive (i.e., homework credit), students may not be motivated enough or feel the need to implement techniques such as successive relearning. Furthermore, without a structured schedule to follow, students may not possess the time-management skills needed to do so. Thus, if educators wish to use successive relearning in their courses, finding ways to implement it into regular course activities may help to ensure learning gains are achieved.

One roadblock that may be encountered when trying to implement successive relearning into courses is that it may not always seem feasible to do so. Admittedly, implementing successive relearning into the biopsychology course in the current study was not an easy task. For instance, developing the materials involved working closely with the instructor to determine which concepts were of vital importance, the extent to which those concepts were represented on the exam, and how to best divide up the concepts into different sets. Furthermore, all materials – including images and diagrams – had to be manually entered into the program using computer software, and students could only access the program by plugging their flash drive into a computer. An important challenge for future research will be to develop an easy-to-use tool (such as an app or online program) that enables successive relearning and that both students and teachers have access to. Another reason successive relearning might be difficult to implement is that it involves revisiting content on multiple occasions, and instructors may not want to spend time during class revisiting material when that time could be devoted to new material. On the flip side of that, instructors may be reluctant to implement successive relearning with out-of-class activities (such as homework) if they do not want to modify the point system for their course or if they fear that students might complain. We should note, however, that in the current experiments students did not seem to mind what was required of them. In fact, several students wrote to us to note how helpful the program was and that they enjoyed using it (for examples, see Appendix B). That students offered these comments voluntarily suggests that they noticed meaningful improvements in their learning, which further testifies to the potential for successive relearning to promote student achievement.

A final issue is the degree to which successive relearning would enhance learning in other content domains or with other materials, because as previously noted, one limitation of the current investigation is that we only evaluated the effectiveness of successive relearning on a single exam. Indeed, our findings demonstrated that successive relearning improved acquisition of relatively complex, science-based material, but the extent to which successive relearning would boost performance in other STEM domains (e.g., chemistry, math) is unknown. Related to this is whether the effects of successive relearning might differ depending on the level of learning tapped by the outcome measure. In the current study, the exam consisted mostly of memory-based questions and had relatively few questions requiring students to make inferences or apply the given concepts (so few that it was not diagnostic/meaningful to analyze them separately). As such, an open questions is whether successive relearning would also lead to a deeper understanding or improve comprehension of material practiced. Preliminary evidence from Rawson et al. (2013) provides some insight into this question. They found that successive relearning did enhance performance on comprehension-based questions, but that this boost was smaller than that observed on memory-based questions. Future work should continue to explore the influence of successive relearning across different levels of learning.

For students to do well in their courses and ultimately achieve in higher education, they need to be aware of and utilize effective learning techniques. We argue that students should treat the learning of course content as they treat the learning of many other skills – that is, by repeatedly practicing that content to mastery. In the current research, doing so produced meaningful improvements in learning as measured by an authentic course exam. Importantly, our findings support the notion that successive relearning is a valuable tool for students and that implementing it into courses may be a promising means of promoting achievement. As with investigations into any new learning technique, future work should continue to examine successive relearning effects across a variety of contexts to establish its boundary conditions and determine when it is most optimal.

References

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, *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.

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.

Dunlosky, J., Hartwig, M. K., Rawson, K. A., & Lipko, A. R. (2011). Improving college students' evaluation of text learning using idea-unit standards. *Quarterly Journal of Experimental Psychology*, *64*(3), 467–484.

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.

McDaniel, M. A. & Little, J. L. (in press). Multiple-choice and short-answer quizzing on equal footing in the classroom: Potential indirect effects of testing. To appear in J. Dunlosky & K. A. Rawson (Eds). Cambridge Handbook of Cognition and Education. Cambridge University Press. NY: NY.

Nevid, J. S., & Gordon, A. J. (2018). Integrated Learning Systems: Is There a Learning Benefit?. *Teaching of Psychology*, *45*(4), 340-345.

Pyc, M. A., & Rawson, K. A. (2011). Costs and benefits of dropout schedules of test–restudy practice: Implications for student learning. *Applied Cognitive Psychology*, *25*(1), 87-95.

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.

Rawson, K. A., & Dunlosky, J. (2013). Relearning attenuates the benefits and costs of spacing. *Journal of Experimental Psychology: General*, *142*(4), 1113.

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.

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*.

Rodgers, J. L., & Shrout, P. E. (2018). Psychology's Replication Crisis as Scientific Opportunity: A Précis for Policymakers. *Policy Insights from the Behavioral and Brain Sciences*, *5*(1), 134–141.

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science, 9*(1), 76-80.

Taraban, R., Maki, W. S., & Rynearson, K. (1999). Measuring study time distributions: Implications for designing computer-based courses. *Behavior Research Methods, Instruments, & Computers*, *31*(2), 263-269.

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.

What Works Clearinghouse (2013). WWC procedures and standards handbook (Version 3.0). Washington, D.C.: Institute of Education Sciences. Retrieved June 28, 2013, from http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19.

Wiseheart, M., Kupper-Tetzel, C. E., Weston, T., Kim, A. S. N., Kapler, I. V., & Foot, V. (in press). Enhancing the quality of student learning using distributed practice. To appear in J. Dunlosky & K. A. Rawson (Eds). Cambridge Handbook of Cognition and Education. Cambridge University Press. NY: NY.

Table 1. *An overview of the course materials used in Experiments 1 and 2.*

Set	Unit	Topic	Stack Name	Practice Type
A	Neuroanatomy	Organization (6)	Midsagittal Structures	Image
		Functions (6)	Midsagittal Functions	Definitions
		Cells in Nervous System (6)	The Neuron	Image
	Neurophysiology	Neurophysiology Terms (6)	Action Potential (AP1)	Definitions
		Neurotransmission (5)	AP Process	Process
B	Neuroanatomy	Organization (6)	Cortex Structure	Image
		Functions (6)	Cortex Functions	Definitions
		Cells in Nervous System (4)	The Synapse	Image
	Neurophysiology	Neurophysiology Terms (6)	Action Potential (AP2)	Definitions
		Neurotransmission (5)	Synaptic Transmission	Process
Base	Neuroanatomy	Organization (9)		
		Functions (9)		
		Cells in Nervous System (4)		
	Neurophysiology	Neurophysiology Terms (6)		
		Neurotransmission (4)		

Note: The numbers in parentheses refer to the number of to-be-learned concepts in that given topic. Base = Baseline. Baseline topics were not available on the flashcard program and so have no stack name or practice type listed. AP = Action Potential. Organization refers to the organization of the nervous system and major brain structures. Functions refer to the functions of major brain structures.

Table 2. *An overview of the homework schedule.*

Date		
E1	E2	Homework Assignment
1/18	8/23	(Flashdrives Distributed)
1/21	8/26	1. Neuroanatomy Stacks (3)
1/24	8/30	2. Neuroanatomy Stacks (3)
1/28	9/2	3. Neuroanatomy Stacks (3)
1/30	9/4	4. Neurophysiology Stacks (2)
2/2	9/7	5. Neurophysiology Stacks (2)
2/5	9/10	6. Neurophysiology Stacks (2)
2/6	9/13	(1st Exam)

Note: Experiment 1 took place in the Spring semester of 2018 and Experiment 2 took place in the Fall semester. The numbers in parentheses refer to the number of stacks assigned for the neuroanatomy and neurophysiology units, each of which corresponded to a topic covered in those units. Across all 6 assignments, students were to complete each of the 5 stacks a total of 3 times.

Table 3. *Student use of the flashcard programs.*

	Experiment 1		**Experiment 2**	
Measure	M	SD	M	SD
Assignments completed	0.86	0.21	0.96	0.10
Assignments completed on time	0.90	0.19	0.74	0.32
Criterion	0.86	0.11	0.88	0.14
Spacing between assignments (days)	3.86	1.76	3.20	2.06

Note. The first two measures are reported as proportions. Criterion = the mean number of times a definition was correctly retrieved during practice.

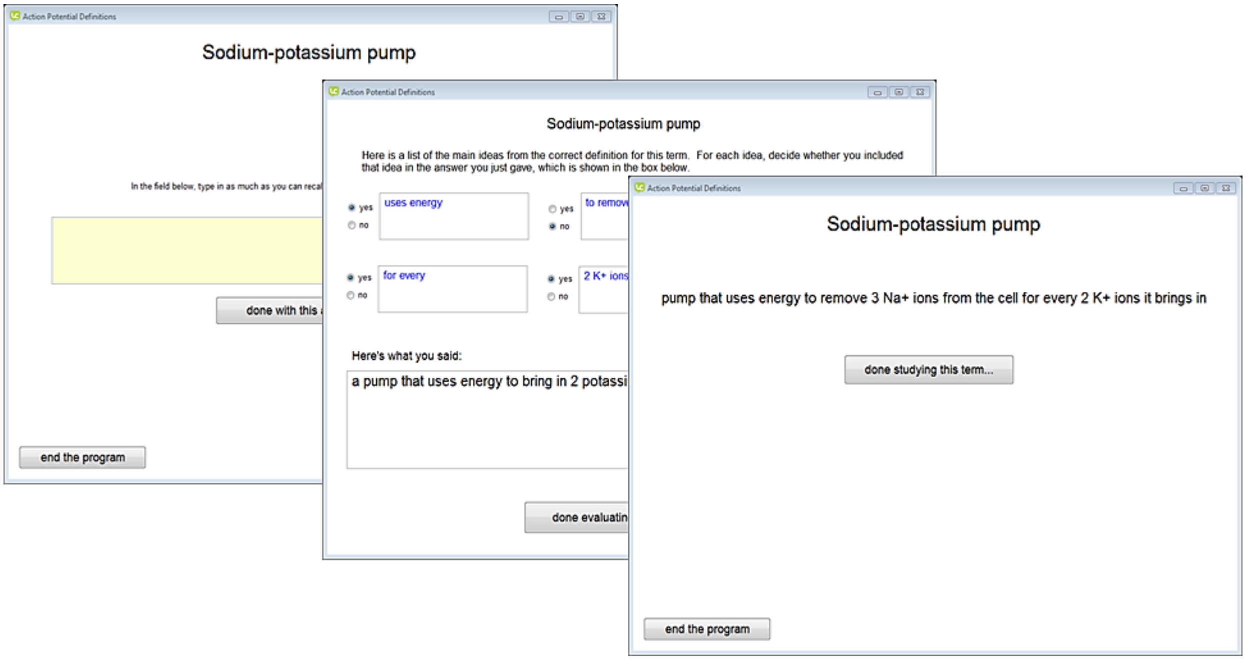

Figure 1. An example of an RMF trial during key-term definition practice.

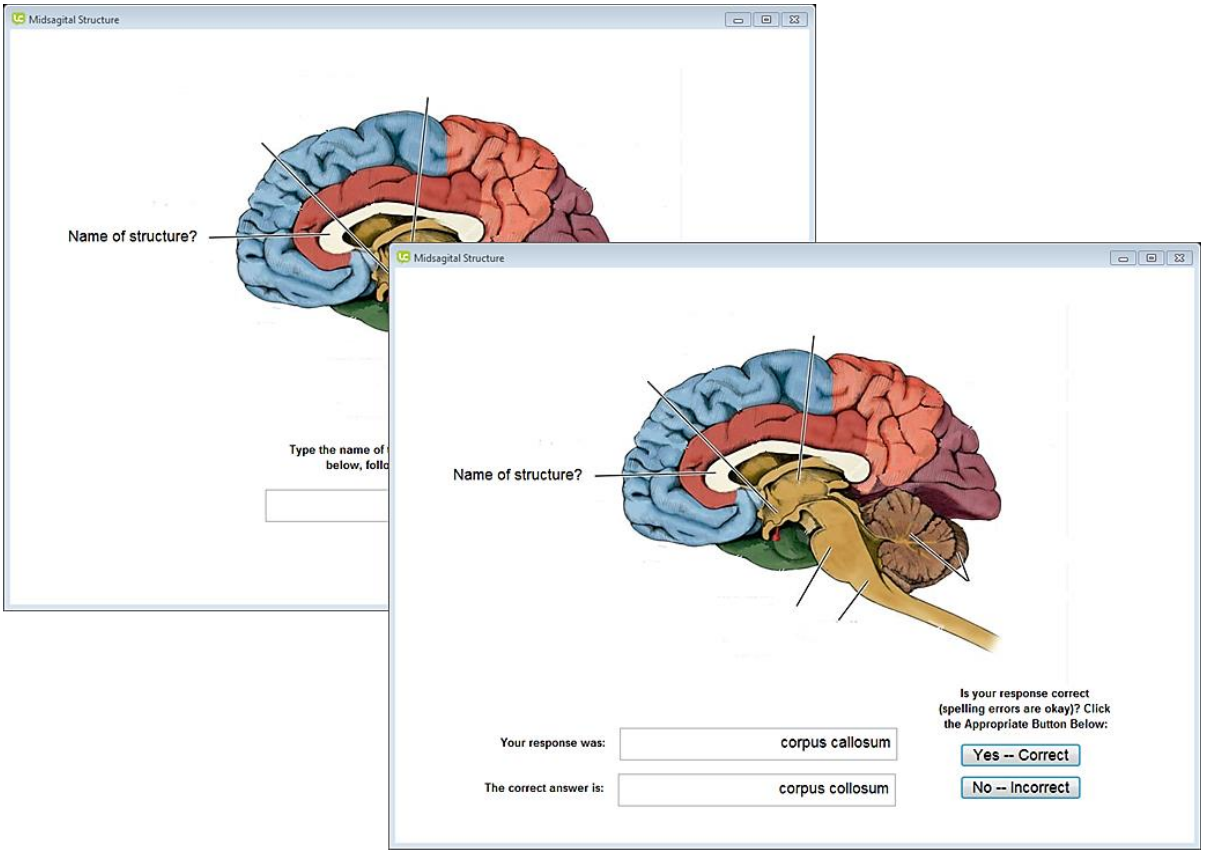

Figure 2. An example of a trial during image practice.

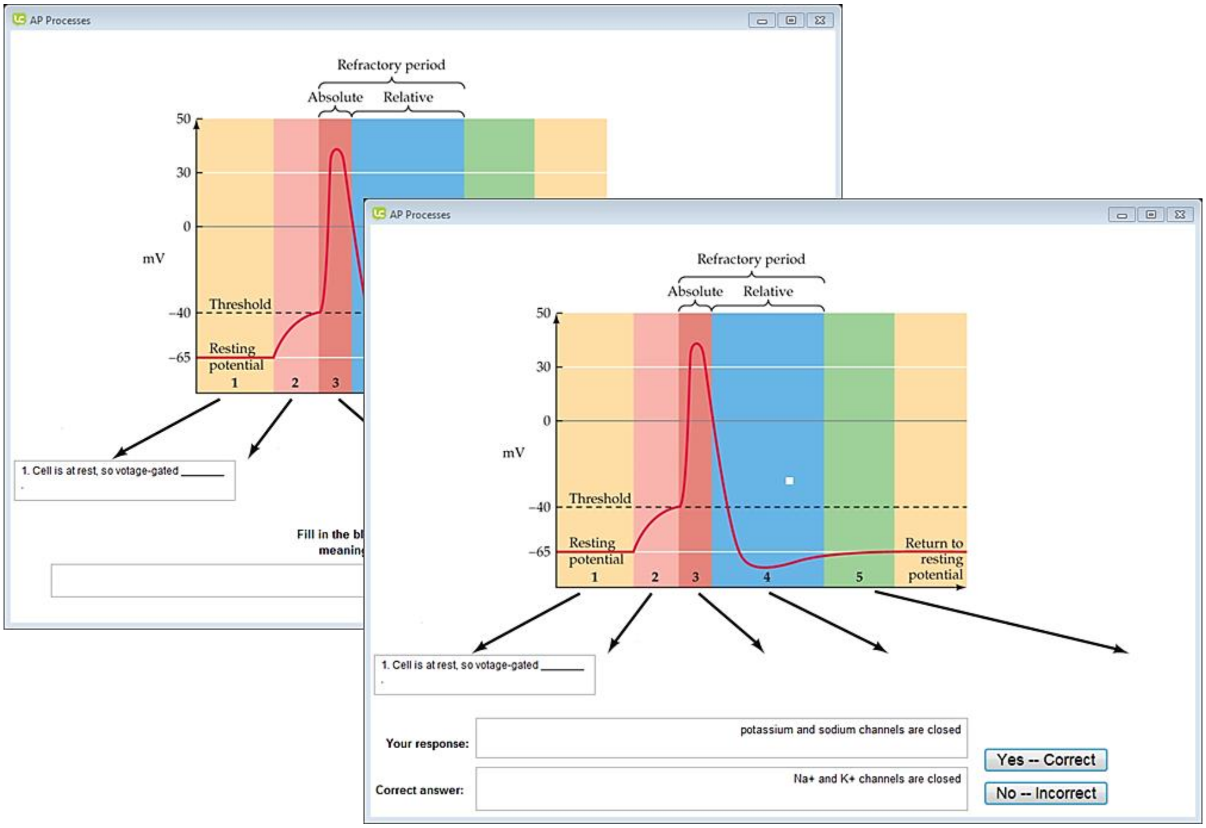

Figure 3. An example of a trial during practice with a process.

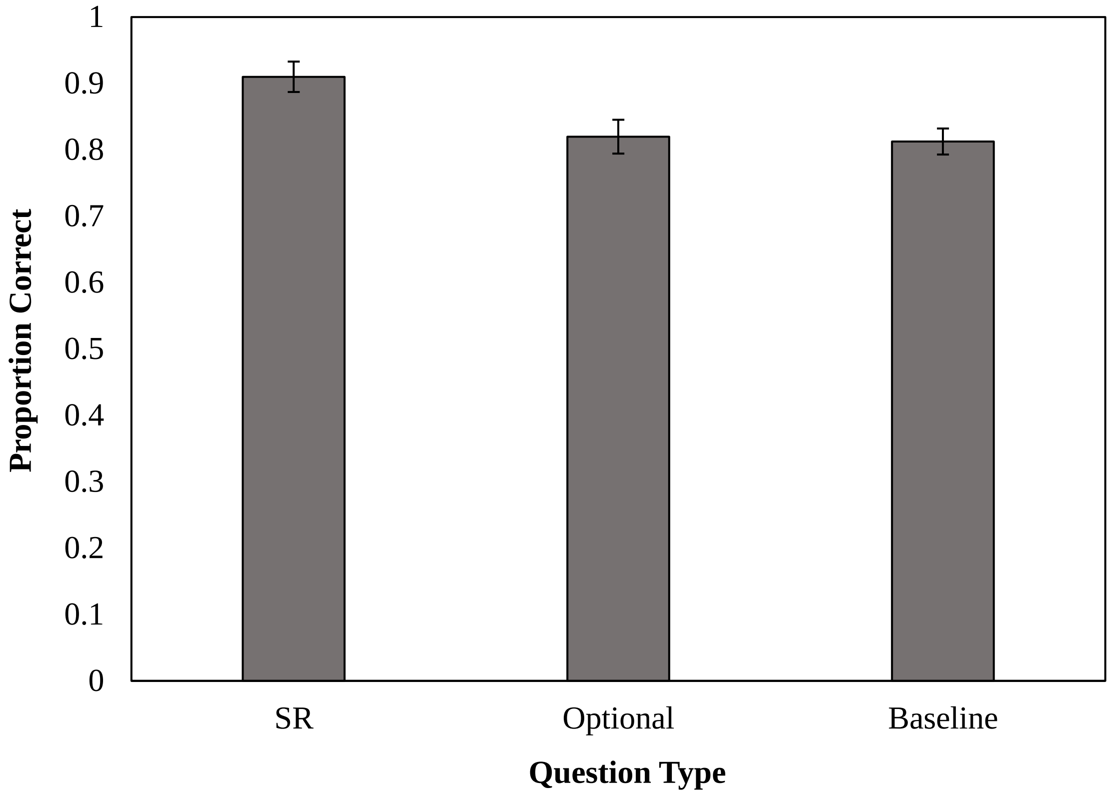

Figure 4. Exam performance broken down by question type in Experiment 1.

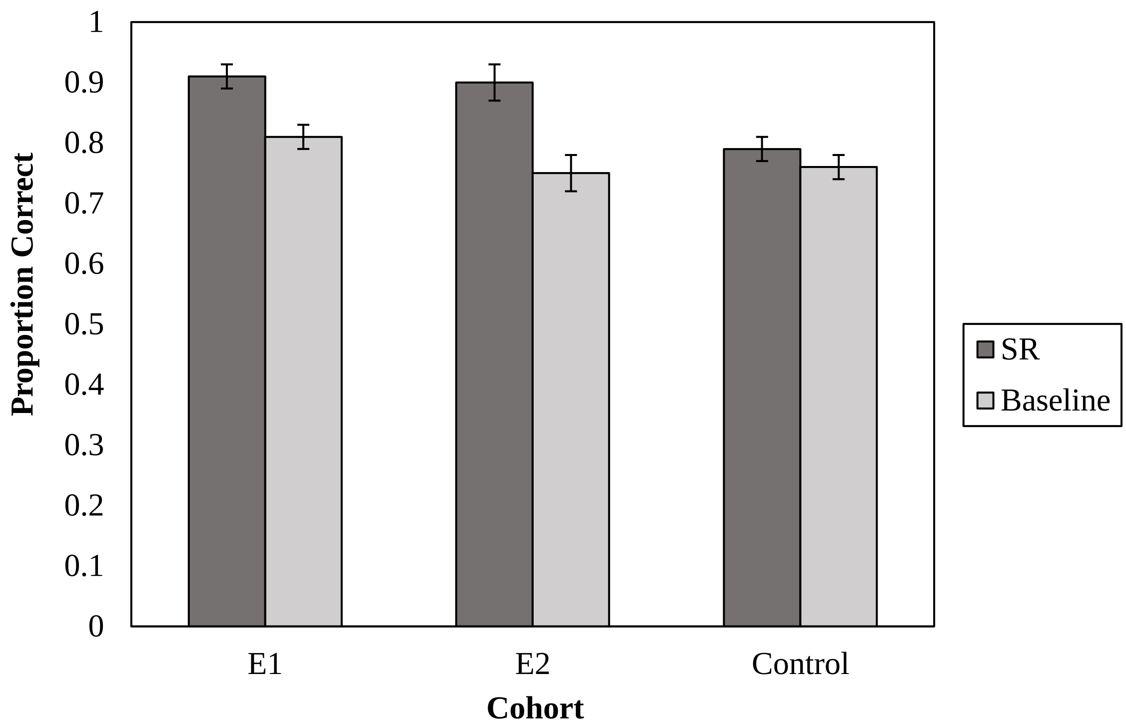

Figure 5. Exam performance for both experimental cohorts compared to a prior cohort.

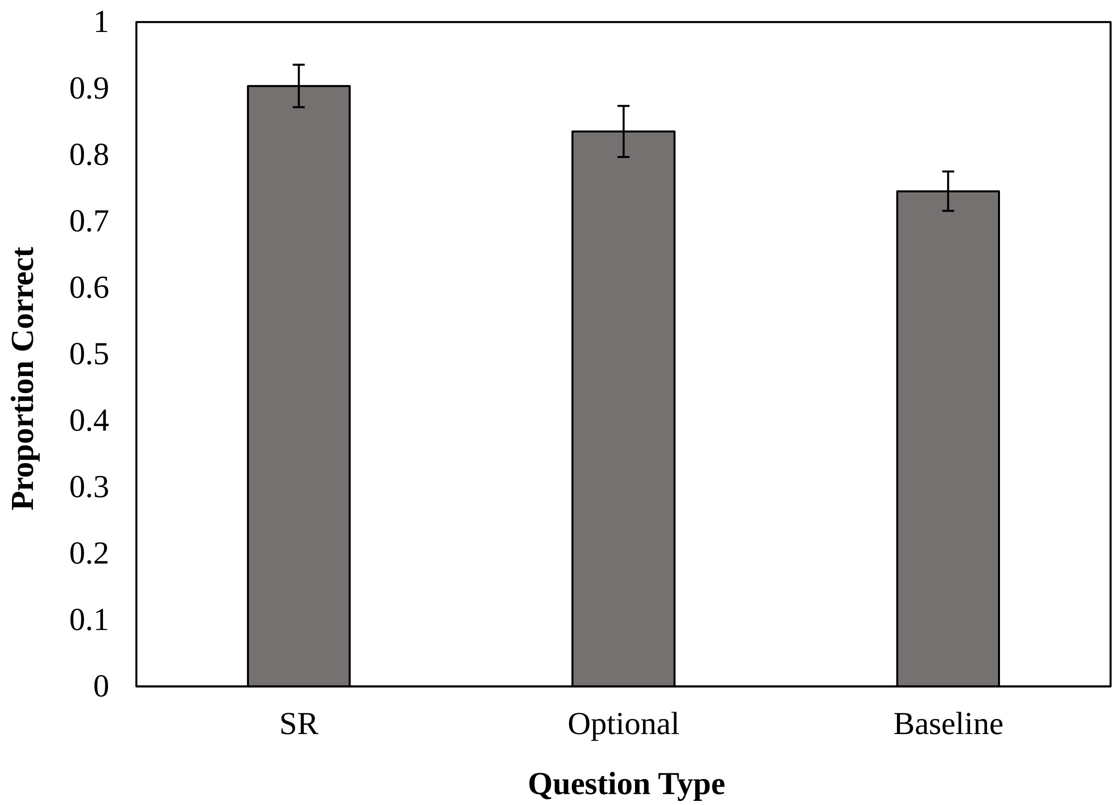

Figure 6. Exam performance broken down by question type in Experiment 2.

Appendix A

Sample Questions from the Exam

A neuron membrane potential moves from -90 mV to -80 mV in response to a brief stimulation. We would term this change in potential as a(n)

a) Depolarization
b) Resting Potential
c) Action Potential
d) Hyperpolarization
e) Inhibitory Local Potential

The Sodium-Potassium Pump is responsible for____________

a) Removing 3 sodium ions from the cell for every 2 potassium ions it brings in
b) Initiating the action potential
c) Depolarizing the cell membrane
d) Returning the membrane potential to the resting state during the action potential

Signals are carried across the synaptic cleft

a) By direct electrical connections between the two cells
b) By secretion of transmitter molecules into the synaptic cleft
c) By transfer of ions from one neuron to the next
d) By carrier molecules
e) By the sodium-potassium pump

The resting membrane potential is

a) negative in neurons and positive in glial cells.
b) partially established by the uneven distribution of ions across the membrane.
c) a property unique to multipolar neurons.
d) established partially by the rapid influx of sodium ions.

The cerebral cortex of the left and right hemispheres are interconnected via axons that travel within the

a) Stria terminalis
b) Cingulate projections
c) Corpus collosum
d) Medial Commissure

Which neural structure below is NOT a part of the limbic system?

a) Cingulate cortex
b) Hippocampus
c) Caudate nucleus
d) Amygdala

Appendix B

Examples of Testimonies from Student Emails

"I found this to be a very useful study tool and I think doing it frequently improved my overall understanding of class material. Thank you for your help!"

"I think this is a wonderful idea for this class. It was a great help."

"I realized that most of the time I devote to studying is actually preparing to study (making notecards, diagrams, quizlets, etc.), not actively studying. I'm sure I get something out of the process through implicit learning, but there is a time cost to this. I was reluctant to think this program would work with biopsych because I've found that making notecards for upper division classes is inefficient; It's just unrealistic due to the vast amount of information and the concept based material. BUT it was very effective because of the idea units and the active retrieval practice. It was a HUGE help and it took less time to retain the same amount of material till mastery"

"Everything worked great! Learned a lot!"

"These homework assignments really helped me in studying for this first exam. I hope that maybe something will be provided for the second exam?"

"I would love it if there was a program like this for the entire class. It really motivated me to keep studying"

"I really enjoyed the program. The modules that were strictly definitions helped me remember definitions on the quizzes, and the other ones helped to solidify the more complex material within this chapter. Something like this would be perfect for our course"

"I do enjoy the program, very easy to work with"

"I already knew from reading a few of Dr. Rawson's articles that it was effective. It was great to see the evidence from research be put into practice and to be able to get the benefits from it. Hopefully, students discover these benefits on their own and choose to utilize the optional notecards."

"Thanks for making this; it really helps"