INCREASING SELF-TESTING BEHAVIOR IN STUDENTS WITH AN ONLINE EDUCATIONAL INTERVENTION

by

ADAM PATRICK YOUNG

B.A., University of Massachusetts, 2011

M.A., University of Massachusetts, 2013

M.A., University of Colorado, 2015

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfilment

of the requirement for the degree of

Doctor of Philosophy

Department of Psychology and Neuroscience

2018

This thesis entitled:
Increasing Self-Testing Behaviors in Students
Using an Online Educational Intervention
Written by Adam Patrick Young
has been approved for
the Department of Psychology and Neuroscience

_____
Dr. Alice F. Healy

_____
Dr. Matt Jones

_____
Dr. Tim Curran

_____
Dr. Mike Mozer

_____
Dr. R. McKell Carter

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB protocol #15-0813

Young, Adam Patrick (Ph.D., Psychology and Neuroscience)

Increasing Self-Testing Behavior in Students Using an Online Educational Intervention

Thesis directed by College Professor of Distinction Dr. Alice F. Healy

Self-testing is one of the most reliable and powerful memory benefits known to cognitive science. Unfortunately, many students omit this strategy in favor of less effective study techniques such as restudying. The superiority of self-testing to restudying is known as the testing effect (e.g., Roediger & Karpicke, 2006a). Applied interventions have been used to attempt to increase self-testing behaviors in student populations, but have so far been unable to draw strong conclusions due to the indirectness of their measurements and incompleteness of their experimental design. Furthermore, these past interventions have been conducted entirely in person, limiting their practical utility relative to the use of an on-demand and widely distributable format. In the present investigation, two experiments were conducted to demonstrate the effectiveness of the Quiz-Me online intervention that seeks to elicit increased use of the testing effect with (Experiment 1) or without (Experiment 2) an in-person lecture component, as well as with (Experiment 2) or without (Experiment 1) random assignment to a control intervention. These experiments also improved upon past research by using more direct assessment of study strategies that students appreciate and use, by assessing individual differences in psychological characteristics, and by establishing benchmark levels of

student conversion from “restudiers” into “self-testers.” Results indicate that in-class lecture components are necessary for eliciting large-magnitude changes to self-testing appreciation and adoption, but that online interventions are nonetheless effective. Despite being smaller in magnitude, they are more strongly associated with changes to testing effect appreciation and changes to strategy use than a control intervention. These findings can be used by instructors and the developers of educational tools to refine the characteristics of future interventions into raising awareness and use of the testing effect among students.

ACKNOWLEDGEMENTS

Thanks to the many people without whom this work would not have been possible. Foremost among them, thanks to Dr. Alice Healy, the most positive, helpful, and genuinely kind person I could have asked for as a mentor. Your guidance, your encouragement, and your trust in me were crucial for my success.

Thanks to my dissertation committee, Drs. Matt Jones, Tim Curran, Mike Mozer, and McKell Carter for their valuable feedback in shaping this research and enhancing its scientific rigor. Thanks also to the many professors and collaborators from the University of Colorado Boulder within the Department of Psychology and Neuroscience, the Institute for Cognitive Science, and the Center for Research on Training for molding me into the researcher I am today.

I'd also like to thank the many members of the University of Colorado Boulder Department of Computer Science. The educational technology described in this thesis is the product of countless hours of design, development, and testing conducted by the following students: Siqi Feng, Bu Sun Kim, Mitch Lewis, Dylan Nguyen, Qi Pei, Scott Richards, Ryan Tabler, Cameron Tierney, Spencer Wilson, and Zichao Yang. Thank you for ingenuity and your persistence. Thanks also to Rick Parker, director of senior undergraduate design projects, for making this great collaboration possible.

Thanks also to the generous funding that has financially supported this project, including the National Science Foundation Division of Research on Learning in Formal and Informal Settings, the University of Colorado Institute of

Cognitive Science Research Grant, and the University of Colorado Graduate School Beverly Sears Research Grant.

I dedicate this thesis to my family: Monica, Calvin, Halle, Charles, and Heather Young. This work was made possible through the strength of their love and support. Thank you!

CONTENTS

TABLES

Table

1. FIGURES

Figure

INTRODUCTION

It should come as no surprise to educational instructors that students often engage in suboptimal learning strategies and exam preparation behaviors. Still, it is one of the great ironies that the people with whom academic researchers and professors interact on a daily basis continue to exhibit a sizeable disconnect between their strategies used and those that have been known from years of research to reliably enhance learning outcomes. To the extent that empirically verified psychological principles go unused, the value of this research is curtailed. The present investigation describes the development and implementation of an educational intervention meant to rectify this unfortunate disparity between the *is* and the *ought* of student learning strategies by increasing the awareness and use of self-testing.

In a thorough and influential review of the psychological research on learning strategies, Dunlosky, Rawson, Marsh, Nathan, and Willingham (2013) identified practice testing (or self-testing, also known as *retrieval practice*), along with distributed practice, as the strategies with the highest utility and educational benefits available to learners. The finding that practice testing is associated with substantially increased learning outcomes relative to other common strategies like restudying was summarized in a meta-analysis by Rowland (2014), who analyzed 159 effect sizes from 61 qualifying studies on the benefits of testing and found a median effect size of $g = .55$, which can be interpreted like Cohen's d and therefore represents a difference of about half a standard deviation in performance between

conditions engaged in practice testing compared to conditions engaged in restudying. A more recent meta-analysis of the benefits of self-testing, this time of 195 relevant studies, found a similar effect size of g = .51 (Adesope, Trevisan, & Sundararajan, 2017). Another recent meta-analysis assessed the benefits of retrieval practice over restudy for the transfer of learning across contexts, and even in these novel situations testing was associated with strong memory benefits and an associated effect size of d = .40 (Pan & Rickard, in press). These reliable benefits of taking practice tests over restudying information is termed *the testing effect* and has been the subject of prolific psychological and educational research (Figure 1).

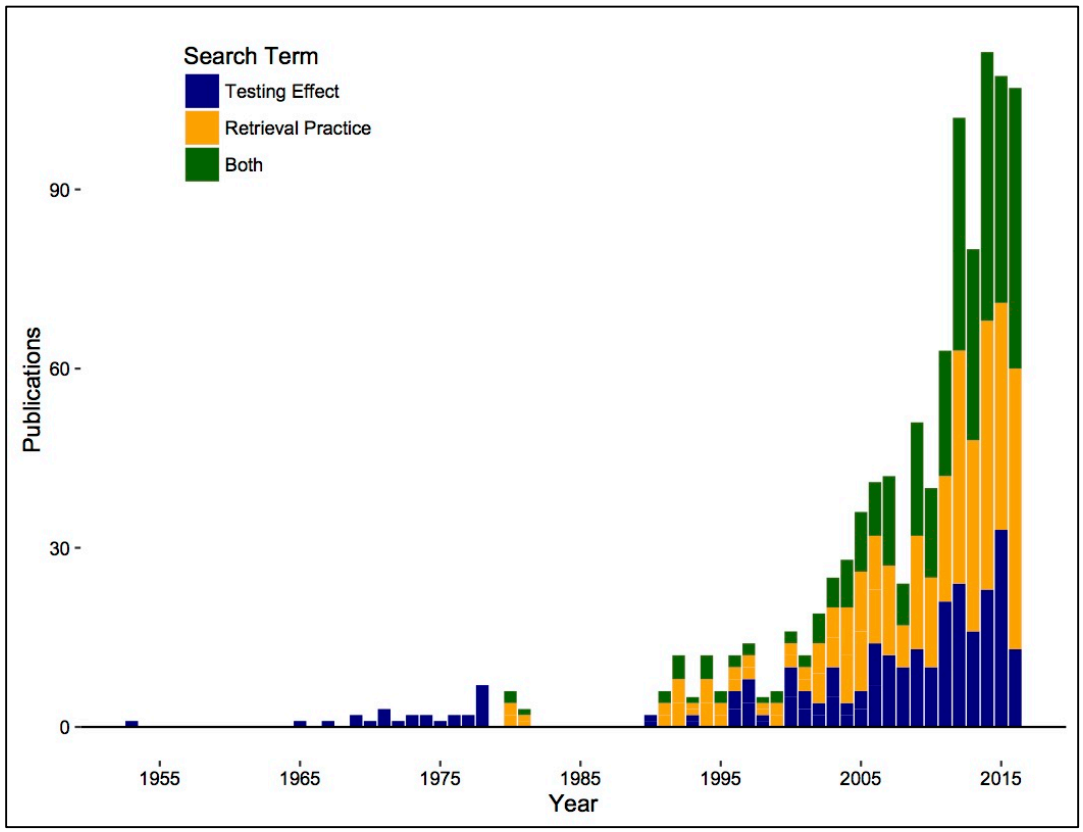


Figure 1. Number of publications returned in Web of Science search categorized by search term including "testing effect," "retrieval practice," or both. Conducted 4/3/17.

Despite its reliable benefits for student learning outcomes, the self-testing strategy often goes underappreciated and underutilized. Karpicke, Butler, and Roediger (2009) found that only 54% of the Washington University sample they surveyed reported solving practice problems or using the self-testing strategy when studying, compared to 84% who said they would re-read their notes or textbook. Moreover, only 13.5% said they used self-testing most often, compared to 54.8% who would re-read their book or notes most often. Even when students do use self-testing, 68% do so for its indirect metacognitive benefits in which they learn which topics deserve more study, whereas only 18% do so intentionally for the direct mnemonic benefits (Kornell & Bjork, 2007). Clearly, student appreciation and use of the testing effect is far below the levels warranted by its documented educational benefits.[1]

Much of the empirical testing effect research has been focused on demonstrating the testing effect across a variety of sample characteristics and procedural details such as materials, test format, and retention interval. Such studies have illustrated the wide applicability of the testing effect in both lab-based and ecologically valid contexts, perhaps most importantly within actual classroom environments. For instance, McDaniel and colleagues (e.g., McDaniel, Roediger, & McDermott, 2007; Roediger, Agarwal, McDaniel, & McDermott, 2011; see also

[1] Recent research has demonstrated that there exists considerable variability in the use of practice problems and self-testing across samples. Taken together, the results of Yan, Thai, and Bjork (2014), Hartwig and Dunlosky (2012), McAndrew, Morrow, Atiyeh, and Pierre (2016), and Geller et al. (2017) demonstrate an average of 68.75% of students reporting the use of practice problems or self-testing, with a standard deviation of 15.65% between studies.

Roediger & Karpicke, 2006b, for a review) have demonstrated that testing effects are reliably elicited in the classroom, thereby emphasizing the educational opportunity that practice tests hold for students using authentic course materials. Despite these advances to our understanding of the reliability and ecological validity of the testing effect, much less research has been conducted in the attempt to empower students with the propensity towards voluntary self-testing.

Only a handful of studies have been conducted that sought to modify students' exam preparation behaviors by teaching them the utility of self-testing. For instance, Young, Healy, Jones, and Curran (2016) sought to determine whether providing brief lectures on the benefits of self-testing and periodic reminders to use practice tests would be effective in increasing self-testing usage and academic performance. To this end, two of four lab sections associated with a college statistics course were provided a 5 min lecture on the testing effect. This lecture consisted of one slide that showed Figure 1 of Roediger and Karpicke (2006a), which was described by the lab instructor, and the following conclusion was emphasized: "Information that was quizzed earlier was remembered far better than information only restudied." Students were then encouraged to take practice quizzes to enhance their memory, and in the week prior to each in-class exam, lab instructors announced the availability of practice tests and encouraged students to take them. Two labs in the control condition were given no such lecture on the testing effect, but were made aware of the availability of practice tests on the course website. Throughout the semester, the 68 students were asked questions in the primary

statistics lecture hall using an iClicker device. Several of these questions were part of the present experiment, such as “While preparing for the last test, which best describes how you made use of the practice test?” Answer options included “I went through and answered questions like I would on a real test,” “I used it like a study guide to identify important concepts and skills for later study,” “I tried to memorize the answers that were provided,” and “I prepared in other ways, without using the practice tests.” Other questions took the form of “For the last test, did you X?” where X cycled through the options “study the notes and lecture slides,” “make up your own practice test and test yourself,” “look over your past homework and lab assignments,” “rework problems from homework and lab assignments,” and “work through problems in the book that were not required for homework.” Response options were Yes, No, and Somewhat. It was hypothesized that the experimental condition (the labs presented with the brief lecture on the testing effect) would demonstrate more behaviors representing self-testing strategies and fewer behaviors representing restudying strategies. Contrary to this hypothesis, a series of chi square analyses revealed no substantial differences in the pattern of student responses by condition. Subsequent analyses further revealed no differences between the groups in terms of exam or course grades. Therefore, the simplest potential solution to the problem of students’ ignorance of the testing effect was not effective in increasing self-testing behaviors. Clearly, different interventions would be required.

If a lecture-based intervention was not effective in raising self-testing rates, what kind of intervention might be more promising? Insight could be potentially be gained from DeWinstanley and Bjork (2004). In their experiments, they succeeded in raising use of the generation effect, a learning strategy similar to the testing effect but which occurs during initial construction of information rather than retrieving information previously encoded (Karpicke & Zaromb, 2010). Their intervention (of a lab rather than a classroom sample) was successful because subjects were exposed to two learning conditions, one that required them to generate information (letters were missing from critical to-be-encoded words), and another that required them to simply read comparable text. After taking a test on this information (without feedback provided), they were given the same two-condition task again, this time with new text content. Whereas the test on the first set of content was associated with a clear benefit of generating rather than reading information, this benefit was eliminated on the second test. This significant condition-by-test-number interaction on test performance was interpreted by the authors as evidence that subjects had become aware of the mnemonic disadvantage of read-only items and then subsequently engaged in enhanced processing of these read-only items for compensation. The results by DeWinstanley and Bjork (2004) thus provide evidence that direct experience with the deficiency of one strategy relative to another might motivate and elicit use of that preferable strategy in the future.

Following this rationale, Young et al. (2016) designed two lab experiments that sought to elicit self-testing behavior after students had experienced both a restudy and a self-testing condition. In these experiments, students were asked to study learning materials (statistics concepts in Experiment 1, n = 106, and chemical element paired-associates in Experiment 2, n = 135), followed by re-studying half of the content and taking a practice quiz on the remaining half. After this review phase, subjects were given a test on what they had learned, and feedback was displayed to indicate whether the respective trial had previously been restudied or practice-tested. This feedback was provided to contrast the memory benefits provided by these two learning strategies and hopefully to make explicit the strategy comparison that was left implicit by DeWinstanley and Bjork (2004). Therefore, unlike the class-based quasi-experiment by Young et al. (2016), the lab experiments provided subjects with evidence that their own performance was superior with self-testing rather than providing evidence taken from a previous study. Following the feedback stage, students were exposed to a new set of learning materials. Experiment 1 used a procedure that largely paralleled that of DeWinstanley and Bjork (2004) by presenting the subjects with both restudy and self-testing trials, and the primary hypothesis was that these subjects would exhibit a condition-by-test-number interaction on test performance. Experiment 2 replaced the half restudy/half practice test phase with a series of screens (one per trial) presenting a previously-studied cue with an instruction to press the spacebar to reveal the corresponding target when ready; in this way, each trial could be either a

restudy or a practice test trial, depending on the learner's preference. The primary hypothesis was that this (experimental) group of subjects would exhibit longer response latency to reveal the target than would a control group who had not been exposed to both restudy and self-testing strategies earlier in the procedure (thereby indicating that the experimental group had learned the value of self-testing indexed by waiting to reveal the answer). Unfortunately, the hypotheses of Experiments 1 and 2 were not supported. These results were attributed to the weak within-subject testing effects observed, likely due to the short retention interval used in these experiments (longer retention intervals are associated with greater magnitude testing effects; Rowland, 2014). Without experiencing large testing effects firsthand, subjects would presumably have no reason to alter their behavior. Thus, these experiments demonstrate the critical importance of eliciting large testing effects in interventions based on firsthand experience.

None of the Young et al. (2016) investigations found an increase in self-testing behaviors. If these approaches were not able to alter students' use of learning strategies, what kind of intervention could? Two relevant intervention studies warrant particular attention. An in-class intervention by Einstein, Mullet, and Harrison (2012) exposed two classes of students enrolled in an advanced psychology course ("Memory and Cognition") to both firsthand experience of the testing effect and a related lecture on the benefits of retrieval practice. The 52 students were given two text passages (on the topics of "the sun" and "sea otters"), and were instructed to read and then re-read one passage (i.e., restudy), and then to

read and retrieve as many details as possible (i.e., self-test) from the other passage. One week later, the students were given a surprise exam on the contents of the passages, tallied their scores, and then listened to a 20-minute lecture on the testing effect. Results indicated that the tested passage was, on average, remembered more accurately than the restudied passage, thus exposing many students to the testing effect. They then completed an assignment on the testing effect relating to their self-testing experience. At the end of the semester, the students were asked to anonymously rate how often they incorporated testing in their reading and studying compared to the beginning of the semester on a scale from 1 (much less) to 5 (much more). Results indicated that students rated themselves as significantly more likely to test themselves when reading and when studying at the end relative to the beginning of the semester. These results constitute the first demonstrated example of students reporting a greater willingness to engage in self-testing following an educational intervention.

A similar study was later conducted by Dobson and Linderholm (2015). In their experiment, 147 students in an anatomy and physiology course were instructed to alternately engage in three encoding strategies while learning physiology passages: read the passage three times (R-R-R), read the passage, freely recall (test on) information from the passage, then reread it (R-T-R), or read then reread the passage while writing down notes (R-R+N). Immediately after turning in these materials, a 30-question multiple-choice exam was administered, containing 10 items experienced using each strategy (with strategy/item combinations

randomized). One week later, the same exam was administered a second time. A week after the delayed assessment, the instructor of the course presented the results of the exams to the class. These results revealed a large within-subject testing effect, such that the tested passage (R-T-R) was remembered more accurately than the restudied (R-R-R) passage[2]. Importantly, the instructor showed the class these results averaged across the class data rather than on a student-by-student basis, thereby exposing all students to results indicating a testing effect. This feedback was followed by a brief lecture on the testing effect. This lecture focused on the use of self-testing during study: Students were advised to read a small section of notes, retrieve as much of that information from memory as possible, review the notes to re-encode forgotten information, and repeat until sufficient performance is achieved (Dobson, personal communication, 2017). This encouragement was repeated leading up to each exam. All of the students who participated to this point were labeled the testing group, to be contrasted with the control group – students in the instructor's previous year's course covering the same material who did not receive exposure to the benefits of testing. Performance did not differ between the experimental and control groups on Exams 1 or 2, as expected, given that the intervention occurred between Exams 2 and 3, but the experimental group outperformed the control group on all subsequent exams. The authors interpreted these findings as an indication that the group exposed to the

[2] Exam performance on the R-R+N strategy was between the R-T-R and R-R-R strategies, but did not significantly differ from either.

testing effect had learned the benefits of self-testing and used this strategy to bolster their performance on subsequent exams.

These two studies demonstrate the promise that might ultimately be attained by educational interventions of the testing effect. Unlike the investigations by Young et al. (2016), both the Einstein et al. (2012) and Dobson and Linderholm (2015) studies made use of the conjoint implementation of direct instruction and comparative experience with positive results, indicating that students can be brought to an increased awareness of the testing effect and use of the self-testing strategy when the intervention contains both of these features. Furthermore, these studies highlight the importance of a lengthy retention interval. Both Einstein et al. and Dobson and Linderholm used a week-long retention interval, whereas Young et al. used retention intervals spanning 5 to 20 min. As students would only be expected to adopt the testing effect after seeing a substantial benefit to their performance by using it, and because longer retention intervals are one of the factors known to be associated with larger testing effects (Roediger & Karpicke, 2006a; Rowland, 2014; for a review of these factors and their implementation in testing effect interventions, see Table 1), using delays of this length is clearly an important factor for incorporation into educational interventions on the testing effect.

Table 1
Factors known to influence the magnitude of the testing effect and their implementation in educational interventions.

Factor	*Rowland (2014) Prescription*	*Einstein, Mullett, & Harrison (2012)*	*Dobson & Linderholm (2015)*	*Quiz-Me.co*
Stimulus Type	Prose or paired assoc.	Prose	Prose	Prose
Initial Test Type	Cued recall	Free recall	Free recall	Cued recall
Retention Interval	≥1 day	1 week	1 week	≥ 1 day
Initial Test Cue-Target Relationship	Semantically related	N/A	N/A	Semantically related
Final Test Type	Cued recall	Cued recall	Cued recall	Cued recall
Initial Test Feedback	Yes	No	Yes	Yes

Despite these virtues, both studies are beset by notable procedural limitations, such as the use of dependent variables that cause issues for interpretation. Einstein et al. (2012) derive their conclusions from measures of self-rated willingness to use retrieval practice, but rather than conducting these assessments in a pre/post-test fashion surrounding their intervention, ratings were measured only at the conclusion of the semester. This procedure means that the judgments that students made with respect to their behaviors at the beginning of the semester could have been inaccurate due to the considerable amount of

intervening time been experience and self-report. If students could clearly recall instances of self-testing a few weeks ago but have difficulty recalling such behaviors that occurred several months earlier, they might mistakenly assume that their behaviors were more drastically different than they actually were. An additional problem for the Einstein et al. data is that the self-testing ratings question involved an explicit comparison of pre-intervention and post-intervention self-testing behavior, so the subjects would have been well-aware of the purpose of this question and thus susceptible to demand characteristics. This susceptibility is especially likely considering that the sample consisted of advanced cognitive psychology students who could be biased towards affirming normative behaviors (such as self-testing) over time, because doing so is generally consistent with the behavior of advanced students. Finally, the self-ratings were significantly different when compared to a model predicting no change from beginning to end of semester (i.e., their mean rating was significantly higher than 3 on the scale from 1-5) rather than compared to data from a control group that did not experience the intervention yet was asked the same rating question, who might have rated themselves as higher than a 3 out of 5 as well for reasons entirely unrelated to an intervention. The joint issues of inaccurate retrospection, demand characteristics, and the lack of a control group make the conclusions by Einstein et al. difficult to accept at face value. Furthermore, Dobson and Linderholm (2015) based their primary conclusions on the inference that the increase in the intervention group's grades over time is attributable to increased self-testing behavior. However, no direct measurement of

self-testing attitudes or behaviors was conducted, so alternative explanations cannot be ruled out. For instance, the classes were taught by one of the authors who was aware of the identity of the control and experimental groups (into which participants were not randomly assigned) and the relevant hypotheses, and even unintentional changes to one's teaching style may have played a role in the different performance levels of the relevant groups. For these reasons, a follow-up study seems justified that (a) measures attitudes towards self-testing and self-testing behaviors rather than course exam data alone, (b) uses a pre/post-intervention design rather than retrospective ratings, (c) includes random assignment to a control group in a double-blind procedure, (d) does not confine its sample to the author's own course.

A follow-up study would also have the additional opportunity of advancing the development of self-testing interventions not restricted to the amelioration of biases and the exclusion of problematic interpretations. One advancement would be to collect data to be used for the sake of benchmarking; if educational interventions are to be designed and implemented, data should be collected that indicate what percent of students benefitted from that intervention rather than concluding that the class has improved as a whole. For example, Einstein et al. (2012) note that 67 and 82% of students rated themselves as somewhat or much more likely to use testing when readying or studying, respectively, compared to the beginning of the semester. These data are problematic due to hindsight bias, and Dobson and Linderholm (2015) do not report data at this level of analysis. Follow-up studies

should report such data to allow for the refinement of educational interventions that improve in efficacy across iterations. Another procedural advancement would be to include individual difference measurements of psychological characteristics in order to identify the predictors of successful or unsuccessful response to the intervention. A final advancement would be to consolidate the intervention into a form that lends itself to a less laborious implementation on the part of the person(s) assigning it. The aforementioned interventions both require instructors to compile relevant materials, allocate class time to the intervention, tally the results, and allocate additional class time to present feedback and advocate for further self-testing. These procedures are likely too burdensome to implement on a large scale, thereby limiting their practical impact. Follow-up studies should therefore seek to improve upon these earlier investigations by providing benchmark data to allow for intervention efficacy analyses over time, by measuring individual differences characteristics and relating them to intervention outcomes, and by packaging the intervention in an on-demand medium.

A final feature of educational interventions worthy of emphasis is the distinction between the awareness of the benefits of using a particular learning strategy and the tendency to put that awareness into action. This distinction is important yet has gone largely overlooked. Do students who fail to respond to the intervention do so because they remain unconvinced of the benefits of self-testing, or are they aware of these benefits yet have difficulty in enacting self-testing behaviors in their own exam preparation routines (e.g., due to lack of motivation)?

Properly understanding the obstacles towards self-testing enactment by dissociating these two issues will facilitate the development of appropriate educational interventions in the future. Einstein et al. (2012) and Dobson and Linderholm (2015) have tacitly assumed that students who become aware of the benefits of self-testing will seize this knowledge and enact greater self-testing behavior as a result. Although some students might be motivated enough to do so, it seems fair to speculate that other students will need the intervention to provide more facilitation in the transition from awareness to action. To this end, the Model of Action Phases (MAP; Bamberg, 2013; Heckhausen & Gollwitzer, 1987), a well-regarded social psychological model of behavioral change, may prove useful for educational interventions.

The process of behavioral change has sometimes been described as a stage-based process. One such stage-based model of behavioral change is the MAP. In this model, the individuals begin in the predecisional stage in which they are aware of possible courses of action but have not made any decision regarding which to pursue. This stage involves the weighing of desirability and feasibility of competing goals and the resulting formation of an intention to pursue a certain goal (a goal intention), such as improving one's study habits or getting better grades. As recommended by intervention studies that failed to yield behavioral change (e.g., Meslot, Gauchet, Allenet, François, & Hagger, 2016), goal intentions should not be assumed to occur or allowed to remain implicit, but instead should be brought to explicit awareness in order to motivate their intention-setting in subsequent stages

of behavioral change. With this goal intention in place, individuals enter the preactional stage in which they weigh the pros and cons of various strategies and settle on a means to achieve their goal. The intention to enact this strategy is called a behavioral intention; for example, to use a self-testing strategy instead of a restudying strategy. This is the only stage of behavioral change that the Young et al. (2016) experiments focused on by lecturing on the benefits of self-testing and demonstrating those benefits firsthand. Crucial to this stage is maximizing the persuasive appeal of the behavior being demonstrated, which can be accomplished for the testing effect by maximizing the magnitude of the testing effect demonstration (such as by following the recommendations of Rowland, 2014, as shown in Table 1). Unfortunately, holding a behavioral intention often fails to lead to the desired behaviors (i.e., the "intention-behavior gap"; Sheeran & Webb, 2016). A hallmark meta-analysis has shown that such behavioral intentions only explain about 28% of the variance in exhibited behavior (Sheeran, 2002). The preactional stage gives way to the actional stage, characterized by the initiation and implementation of the chosen behavioral strategies, a process facilitated by the use of implementation intentions, if-then rules that individuals establish to guide their behavior in the future. For example, a student may form the implementation intention: "If I sit down to study with my book open, then I will first turn to the chapter questions in order to quiz myself." Implementation intentions are associated with a medium-strong effect size on behavior relative to not using them (Gollwitzer & Sheeran, 2006). Following the use of such implementation intentions,

the fourth and final postactional stage is entered, in which the individual evaluates the outcomes of the behavioral change process, refining the effectiveness of decisions made at previous stages, and avoiding relapse into suboptimal behaviors. Interventions that accommodate these four stages of behavioral change might increase the chances that newfound awareness of the testing effect will lead to the adoption of self-testing behaviors.

The present investigation is an attempt to improve upon previous testing effect interventions in a variety of ways. First, our intervention (hereafter referred to as "Quiz-Me") will measure attitudes towards self-testing, self-testing behaviors, course exam data, and individual difference characteristics. Second, Quiz-Me will derive these self-testing measures from both a randomly-assigned experimental and a control group before and after the educational intervention, and will do so across a large and varied sample using a double-blind procedure to minimize the possibility of biases. Third, Quiz-Me will utilize the Model of Action Phases to facilitate behavioral enactment such as by eliciting goal and implementation intentions. Finally, Quiz-Me will be conducted in a manner that emphasizes practicality of dissemination; this intervention will be hosted online (at www.quiz-me.co) in order to allow the intervention to be conducted at the students' convenience and not during valuable class time. These procedural improvements will be conducted to supplement the best elements retained from previous interventions, such as the

joint action of direct experience and lecture[3] on testing's benefits as well as the use of a substantial retention interval and other elements advised to maximize the magnitude of the testing effect. In doing so, Quiz-Me might prove to be the most effective and practical intervention yet designed for increasing awareness and use of the testing effect.

The present investigation describes two experiments that were conducted to evaluate the effectiveness of the Quiz-Me educational intervention. In Experiment 1, the intervention was assigned to one class and all subjects participated in the experimental condition. In Experiment 2, the intervention was assigned to seven college classes, and students were randomly assigned to experimental or control conditions. We discuss each experiment, and its effectiveness for eliciting self-testing awareness and behavior, in turn.

[3] Lecture is present in Experiment 1 but absent in Experiment 2 to assess its importance for effects observed in Experiment 1.

EXPERIMENT ONE

Experiment 1 was designed to evaluate the effectiveness of a preliminary version of the Quiz-Me educational intervention. Using a sample of one college class in size, all participants were exposed to the experimental version of the intervention (i.e., there was no control condition for this study). This procedure was selected to maximize exposure of participants to the crucial intervention condition given the small sample available and to determine if the intervention would lead to increased testing effect awareness and self-testing behaviors, thus paving the way for a more elaborate follow-up experiment.

Method

Participants

The Quiz-Me intervention was announced as an extra credit homework assignment in a mid-level psychology course ("Introduction to Cognitive Psychology") taught by the first author. On the educational website (to be described below), students were asked for consent to collection of website data and relevant course data. It was made clear that consent was voluntary and the instructor would remain unaware of which students consented for data collection and who participated while declining consent for data collection (yet still earned course credit). An alternative assignment (i.e., reading and summarizing an article on the testing effect) was also offered in lieu of participation, but no students selected this option. The sample consisted of 27 students who consented to data collection and provided complete sets of data. In the interest of maintaining brevity within the

online intervention, demographic data were not requested. Nevertheless, the sample consisted of a typical second-to-third year psychology class at the University of Colorado Boulder: majority female, majority white, majority late teens to early twenties.

Materials

Motivated Strategies for Learning Questionnaire. During the first week of the semester, the class was asked to complete the Motivated Strategies for Learning Questionnaire (MSLQ; Duncan & McKeachie, 2005; Pintrich & DeGroot, 1990; Pintrich, Smith, Garcia, & McKeachie, 1993) on the course website. This survey was administered in order to perform exploratory research of individual differences related to attitudes towards learning (e.g., control over one's learning, effort regulation, metacognition) that might predict positive responses to the Quiz-Me intervention. Original questions were created to assess motivation, self-testing proclivity, and embrace of technology, and these items were randomly interspersed within the survey (see Appendix B). Despite including original items within it, this survey will nevertheless be referred to as the MSLQ for simplicity. All items were scored from 1-7, with 7 representing "very true of me" and 1 representing "not at all true of me."

Study Strategy Survey. The appreciation of the testing effect and use of the self-testing strategy was measured using the Study Strategy Survey (SSS). This survey, shown in Appendix C, displayed a list of eight common study strategies, populated by student data provided in Karpicke et al. (2009). The first three

questions on the SSS asked students to rank these eight strategies (more precisely, only to rank the top 5 for each question) with respect to each question's prompt. Question 1 asked subjects to "Think of the strategies you used to prepare for this test. Rank the top 5 strategies in order of how often you used them, with 1 as most often." Question 2 asked subjects: "Out of all the study strategies (A-H), which do you think lead to the best learning outcomes for the average student? Rank the top 5 strategies, with 1 as best." Question 3 asked subjects: "Out of all the study strategies (A-H), which do you think lead to the best learning outcomes for you? Rank the top 5 strategies, with 1 as best." Question 4 asked subjects: "How beneficial do you believe the *Rereading notes or textbook* strategy is for your memory?." Question 5 asked subjects: "How beneficial do you believe the *Do practice problems, self-tests, or flashcards* strategy is for your memory?." For both Questions 4 and 5, a 1-7 scale was provided, with 1 associated with "Not at all effective" and 7 associated with "Very effective."

Quiz-Me.co Web-Based Intervention. The Quiz-Me intervention was hosted on the website http://quiz-me.co. Screenshots of the website are included in Appendix A. Students were told to access the website only from a desktop or laptop computer, and not from a smartphone, tablet, or other mobile device. Further details of the web-based intervention are described in the Procedure section below.

Procedure

The general procedure involved assigning the SSS as a pre-test along with an in-class exam (on Oct 4th). Two weeks later (on Oct 21st), the online participation component was announced and made available to the class. Four days later (On October 25th), the SSS was assigned as a post-test along with the next in-class exam. Therefore, the procedure progressed from in-class, to online, to in-class again.

In-Class Procedure. The in-class procedure was identical for both pre-test and post-test administrations of the SSS. Paper-based class exams were distributed at the start of the 50 min class period, with the SSS stapled to the final page. Completing the SSS and turning it in along with the exam earned students two extra points on their 100-point exams. All students turned in completed surveys with their exams prior to leaving the classroom.

Online Procedure. Two and a half weeks after this exam (the second exam of the course), the instructor made a verbal announcement in class about the availability of an extra-credit assignment that could be conducted online. This announcement was also delivered via email to the class roster. Students were instructed to go to www.Quiz-Me.co and complete the experience (labeled "Tutorial" at the time) to earn additional extra-credit points. The announcement was made as follows:

> *If you'd like to earn extra credit, please take the Tutorial experience at www.Quiz-Me.co , an in-progress educational website aimed at improving your study strategies. Take this website on a laptop*

> *or desktop computer, not a mobile phone. Also, it does not work properly on Safari [web browser]. The Tutorial takes about 20-30 minutes on day one. 24 hours later, you will receive an email asking you to return to the site and finish. This second session takes about 10-15 minutes. You must complete both day one and day two to earn extra credit. Follow the instructions on the website, and ignore the typos! Importantly, you must finish the assignment by Monday, meaning you must begin the assignment by Sunday at the latest, preferably Saturday. If you would prefer, an alternative assignment is available. Read the attached journal article [Roediger & Karpicke, 2006a] and write a two-page paper summarizing the results and relating it to your own life and study habits. This assignment is also due on Monday. You may do either the paper option or the website option, but not both.*

No students opted to complete the alternative, paper-based assignment.

Unlike on a typical website made for casual browsing with widely flexible navigation options, the online intervention was highly controlled, with a mostly linear procedure. Students were required to begin by clicking "Begin Day 1" at the top of the Home page, and then entering their college-affiliated email address to log in. On the next page, they were shown a consent form and asked to indicate whether they consented to data collection both of their online participation data and of relevant course data. Students could participate for extra credit even if they selected "No" on the consent form, although no students declined consent.

Following consent authorization, students were asked to provide responses to the following three questions, with Options 1 (Strongly Disagree) through 7 (Strongly Agree) displayed: "Being a good learner is important to me," "Receiving good grades is important to me," and "Saving time by studying is important to me." The purpose of these questions was not to measure students' attitudes but rather to get them to think about these values prior to beginning the intervention. In line with the tenets of the Model of Action Phases (Bamberg, 2013; Heckhausen & Gollwitzer, 1987), these value-related thoughts were assumed to evoke relevant goal intentions, presumably priming students (or otherwise subtly preparing them) to respond to the intervention in a goal-conducive manner.

On the next screen, subjects were given instructions for the experiment. Specifically, they were told that today's participation would involve the completion of three tasks: Studying 20 trivia facts, restudying half of them, and taking a practice quiz on the other half. They were also told that a Final Test would occur the next day, which would allow them to see which strategy leads to better memory: restudying, or taking a practice quiz. Students were then instructed to "Pick a topic that you are interested in but do not know much about." Options included: Sports and Gaming, Animals and Biology, Food and Drink, World History, or Anatomy and Medicine. Each category contained 20 true trivia facts, 1-2 sentences in length, and each fact was judged by the author to be esoteric enough that they would be unlikely to be known by non-experts.[4] This procedure was used so that subjects

[4] Due to technical error, students' selection of topic of study was not recorded.

would be interested in the material and less likely to exhibit the passive, unmotivated behavior typical of bored subjects while traversing the online experience.

Study phase 1. After selecting their topic of choice, the initial study phase would begin. Instructions stated that the subjects should "Study 20 facts about the topic of your choice. These facts will be on the Final Test tomorrow, so try hard to memorize them." After clicking "Begin," the first fact, along with a counter (e.g., Fact 1 of 20) would be displayed. Aside from reading the trivia fact in the middle of the screen, the only other action the page allowed was to select "Continue." All 20 facts for that topic would cycle through the page in the same format before concluding the initial study period.

Restudy phase. Students then entered the restudy period, beginning with instructions that stated: "Restudy 10 of the facts you previously studied. These facts will be on the Final Test tomorrow, so try hard to memorize them." After clicking "Next," students would be exposed to 10 sequential screens, each containing one of the 20 previously-studied trivial facts. The only difference between this section and the Study Phase section is that for each fact, one keyword has been underlined. The reason for underlining a keyword was to attempt to control for attentional effects on memory, as practice test trials (described below) clearly direct attention toward the keyword being retrieved from memory. By underlining a keyword during restudy, the differences between conditions is isolated to the processing of the information rather than the distinctness of that information relative to the text around it (for

more on the control of attention in educational testing effect studies, see Shapiro & Gordon, 2012).

Practice quiz phase. At the conclusion of the Restudy Phase with the 10th restudied item, students were given instructions for the Practice Quiz Phase.[5] These instructions stated: "You will now take a practice quiz on 10 of the facts you previously studied. These facts will be on the Final Test tomorrow, so try hard to memorize them." Below these instructions were three examples of the feedback they would receive in this phase. Specifically, an example of a red feedback box was provided and was said to indicate an incorrect answer, but that the user should "check the hint at the bottom and try again." It was also noted that the feedback screen would display the answer that the user typed, and when incorrect, would provide a hint by showing the first letter of the correct answer. By providing immediate feedback and the opportunity to self-correct, it was expected that the magnitude of the testing effect would be larger than without this feature, as consistent feedback has been associated with larger testing effect sizes (Rowland, 2014), and getting a practice quiz correct has been associated with larger memory gains than when getting a practice quiz incorrect (Kornell, Bjork, & Garcia, 2011). If a user got the second attempt incorrect, they would be shown the correct answer. An example of a yellow feedback box was also displayed and was said to indicate a correct answer that had been spelled incorrectly; both the user's spelling and the

[5] The restudy phase was always presented before the practice quiz phase due to the recommendation by Einstein et al. (2012) in order to reduce the possibility that students will engage in covert self-testing during the restudy phase.

correct spelling were provided. Finally, an example of a green feedback box was provided and said to indicate a correct answer. In keeping with the recommendations for optimal feedback by Tack, Healy, and Jones (2014), the correct answer was always provided with the second incorrect attempt or with the correct attempt.

Following these instructions and feedback examples, subjects were exposed to a sequence of 10 practice quiz trials, one per screen. The trials consisted of sentences, identical to those previously studied, but with one keyword missing. For example, a practice quiz trial might read: "Without access to written language, the Inca used a system called [Fill in the blank][6] to communicate using knots in rope or string." The subjects were expected to type their response into that blank, and then press the "Check" button, at which point the appropriate feedback message would be displayed, and following an incorrect response, a second attempt was also permitted. After 10 practice quiz trials had concluded, a message informed the subjects that "Today's participation is complete. In 24 hours you will receive an email with a link asking you to return to Quiz-Me to finish participation. You must finish the Day 2 procedure to receive credit – If you do not finish Day 2, you will receive no credit!"

Final test. After 24 hours, students received an email that said: "Hey [user's first name]!! Come back to Quiz-Me.co to finish participation and receive your extra credit! Enter your email: [user's email address] into the box on the Day 2 Return

[6] The correct answer is given at the end of the next footnote.

page to complete the Quiz-Me experience!!! quiz-me.co/#/returnPage." After returning to the Quiz-Me website, students entered their university-affiliated email address to log back in. Once logged in, they were shown a screen of instructions that read: "Yesterday you studied 20 facts, restudied half of them, and then took a practice quiz on the other half. Now you will take a Final Test on all the facts. It is important that you try to do as well as you can. After the test, you will be shown: 1) How you did on test questions for facts that you Restudied yesterday, 2) How you did on test questions that you Took a Practice Quiz on yesterday… So you can compare which strategy leads to better memory!" After pressing "Start," the Final Test began. This test was identical in format to the Practice Quiz section from Day 1, except there were questions for all 20 facts, not just half of them. Incorrect responses were still followed by second chance attempts. Scoring was such that all correct attempts, regardless of spelling accuracy and whether the correct answer was achieved on the first or second attempt, were graded as "1" for that trial; only when subjects got the answer wrong on both attempts did they receive a "0" for that trial. The Final Test concluded after the 20th trial was submitted.

Feedback page. Immediately after the conclusion of the Final Test, subjects saw the Feedback Page[7] in which their performance on the Final Test was revealed to them. Importantly, the purpose of this page was to elicit in the subject the understanding that self-testing leads to superior learning outcomes than does

[7] This Feedback Page was not as extensive as that shown in Appendix A (used in Experiment 2). The primary difference is that, due to technical constraints at the time, the Feedback Page of Experiment 1 did not shown data relevant to the subjects' own performance, only the performance of the average user.

restudying, and to elicit implementation intentions consistent with enacting self-testing behaviors. First, the top of the Feedback Page displayed a text section that read "How do you think you did? If you are like most people, you did better on Tested (fill-in-the-blank) than Restudied items! The data below show typical results, with a clear advantage for items that were practice tested (data from Rohrer, Taylor, & Sholar, 2010)." Beneath this text was a bar graph that displayed test accuracy on the Y axis and two bars on the X axis corresponding to Practice Tested and Restudied trials. The bars represented a 58% to 42% accuracy advantage for tested over restudied items. True to the text, these data had been taken from Rohrer et al. (2010, their Figure 6) simply due to the impressive yet fairly typical effect size they demonstrated. Underneath this bar graph was another text section with the large header that read "How to Quiz Yourself." Below read the following text:

There are three popular ways to give yourself practice quizzes:

1) *Make flashcards either by hand or online (such as on Quizlet.com), and don't turn them over until you've tried to recall them from memory.*
2) *Think of a topic you think is important to the class and try to write down as much information about it as you can. Then, check your book and notes to see how much you remembered and what you forgot.*

3) *Have a friend ask you questions. Answer them as thoroughly as you can, and have them tell you the information that you forgot to include.*

Beneath this "How to Quiz Yourself" text was a series of instructions designed to elicit implementation intentions: "Think about which strategy you can see yourself using. Research shows that it helps to state your new plan in If-Then format. For example, If I sit down to read my notes, Then I will go to Quizlet.com and make flashcards instead." Following these instructions were a series of prompts with text-entry boxes for the subject to type into, designed to elicit implementation intentions.[8] The first prompt stated: "Write your new plan to quiz yourself below." The word "If:" was shown above the first text box, and the word "Then:" was shown above the second. Below these text-entry boxes was a button labelled "Finish." After progressing past the Feedback Page, subjects were shown the following message: "Your participation in the Quiz-Me experience is now complete, and your instructor will be notified that you have earned participation credit." The following week, the students took the next in-class exam and Study Strategy Survey (see In-Class Procedure section), which completed their participation in the study.

Results

A total of 27 subjects provided consent and a complete set of data (i.e., they did not miss either in-class exam nor fail to complete the Quiz-Me intervention).

---

[8] The conditional *if-then* nature of implementation intentions has been found to be particularly important, as they lead to stronger behavioral consequences than plans to act that take an *I will – when* format (McCrea, Penningroth, & Radakovich, 2015; Pakpour et al., 2016).

Occasionally, subjects failed to respond appropriately to the ranking questions, such as by listing several strategies with the same ranked numerical priority (e.g., Restudy and Self-Testing each given the rank of 1). Whenever these invalid rankings were observed (for one subject on the Pre-Test and one subject on the Post-Test), rankings of the same numeral were kept equally prioritized, but the best ranking (i.e., the smaller numeral) possible for subsequent strategies was altered to avoid rank-inflation (see Table 2 for a representative example of this recoding policy). This policy was adopted to maintain the subject's representation of relative rankings but reduces artificial rank inflation (or rather deflation, given lower ranks' higher priorities). Most subjects followed instructions and only provided ranks for positions 1 through 5; some subjects, however, ranked all 8 strategies. For the sake of consistency, all unranked strategies and strategies ranked 6-8 were recoded into a rank of 6, thereby sharing the last place priority spot. This policy was adopted so that the strategies' mean rankings would not be influenced by the presence of rankings of 7 and 8 from some subjects.

Table 2
An invalid Study Strategy Survey response and its corresponding values after recoding

	A	B	C	D	E	F	G	H
Original	1	1	2	2	3	4	5	6
Recoded	1	1	3	3	5	6	6	6

Hypothesis 1: Increased Testing Effect Appreciation

It was hypothesized that the Quiz-Me intervention would cause students to gain appreciation for the testing effect. Therefore, it was predicted that the students' average rank of response option B ("Do practice problems, self-tests, or flashcards") would be significantly lower (i.e., more prioritized) after the intervention than before it. Similarly, it was predicted that the average rank of response option A ("Rereading notes or textbook") would be significantly higher (i.e., less prioritized) after the intervention than before it. These predictions were made for both Study Strategy Survey Question #2 ("Which do you think lead to the best learning outcomes for the average student?") and Question #3 ("Which do you think lead to the best learning outcomes for you?"). It also predicted that SSS Question #5 ("How beneficial do you believe [practice testing] is for your memory?") would be significantly higher after the intervention than before it, and that SSS Question #4 ("How beneficial do you believe [restudying] is for your memory?") would be significantly lower after the intervention than before it. Response patterns for restudying and self-testing strategies on Question 2 and 3 are provided in Figures 2 and 3, respectively.

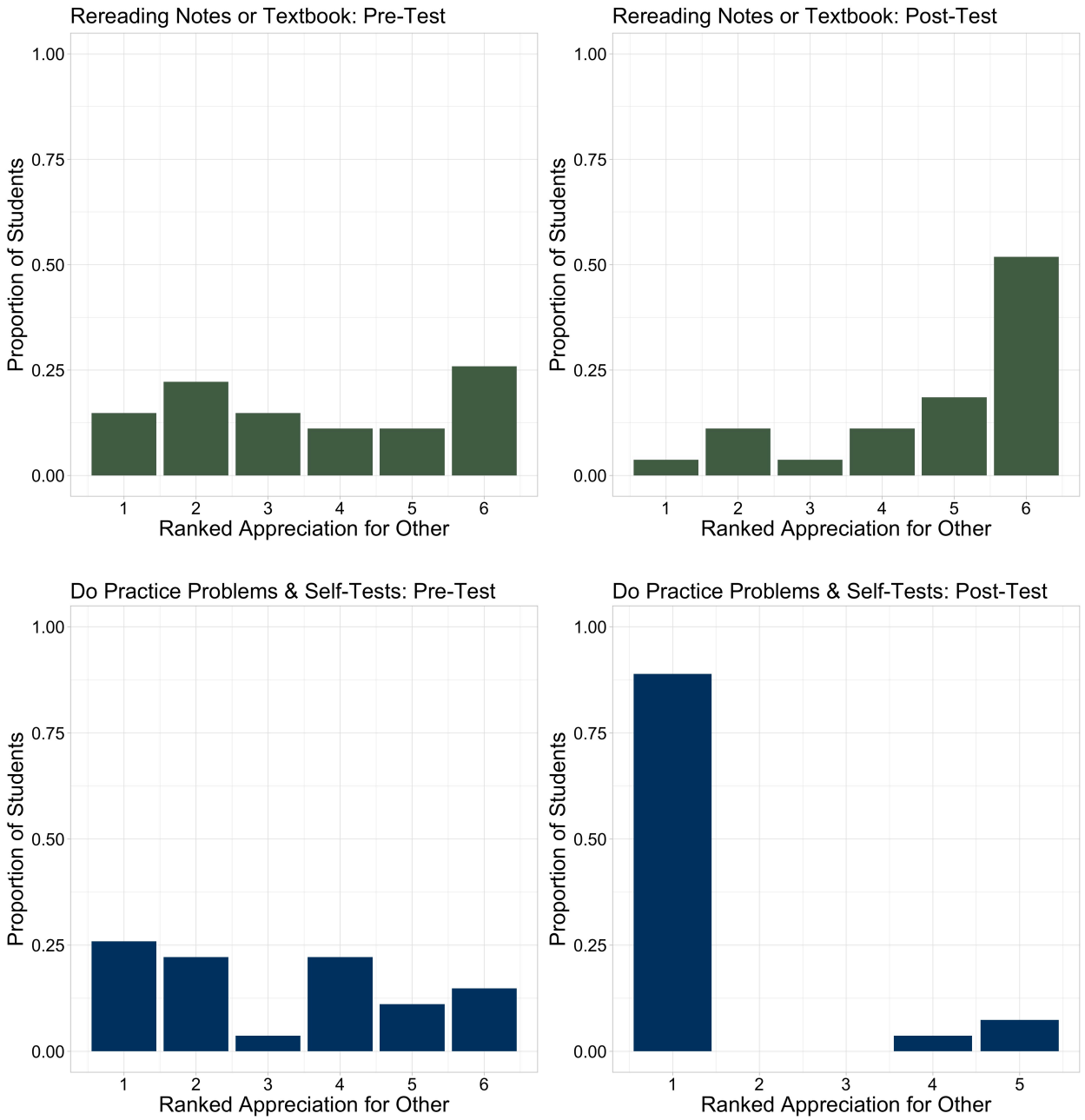


Figure 2. Study Strategy Survey responses for Experiment 1: Question #2

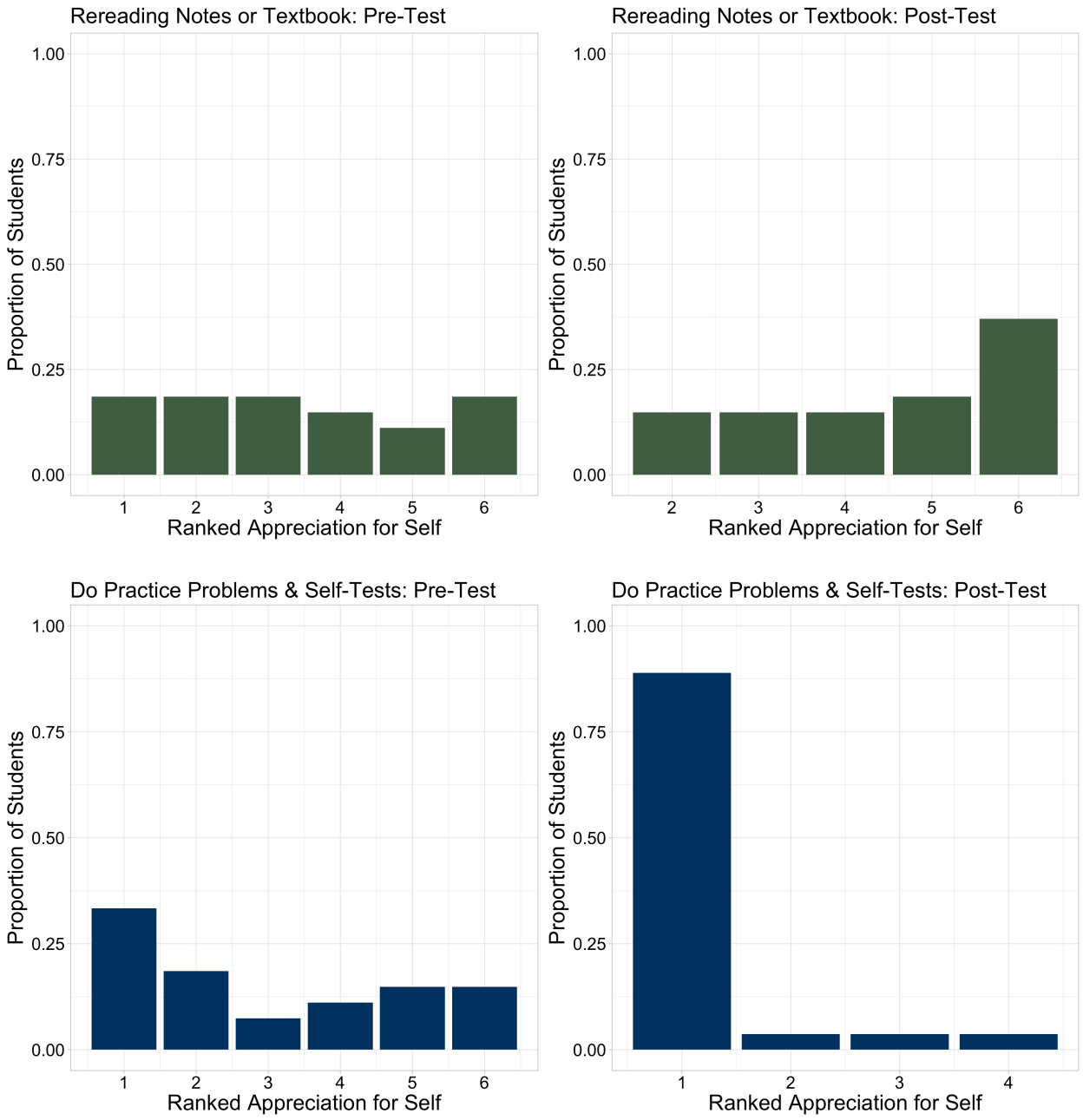


Figure 3. Study Strategy Survey responses for Experiment 1: Question #3

For SSS Question #2, response option B (self-testing) exhibited a significant decrease in rank from pre-test to post-test, t (26) = 4.497, p <.001, d = .865, a large effect. Concurrently, response option A (restudying) exhibited a significant increase in rank from pre-test to post-test, t (26) = 3.164, p <.01, d = .608, a medium effect. Surprisingly, response option C ("Rewriting notes") also exhibited a significant increase in rank from pre-test to post-test, t (26) = 2.935, p <.01, d = .564, a medium effect. Prior to the intervention, 48.14% of students reported more appreciation (i.e., gave a lower, more prioritized rank) for self-testing's normative benefit than for restudying, but after the intervention, 92.59% did so, a marginally significant increase, χ^2 (1) = 3.789, p = .051.

For SSS Question #3, response option B (self-testing) exhibited a significant decrease in rank from pre-test to post-test, t (26) = 4.742, p <.001, d = .912, a large effect. Concurrently, response option A (restudying) exhibited a significant increase in rank from pre-test to post-test, t (26) = 3.126, p <.01, d = .601, a medium effect. Prior to the intervention, 48.14% of students reported more appreciation (i.e., gave a lower, more prioritized rank) for self-testing's personal benefit than for restudying, but after the intervention, 100% did so, a significant increase, χ^2 (1) = 4.900, p < .05.For SSS Question #4 and 5, responses[9] representing the perceived magnitude of the mnemonic benefits of self-testing did not significantly increase from pre-test (M = 6.481, s = 1.051) to post-test (M = 6.703, s = 0.608), t (26) = 1.236, p > .05, d = .238,

[9] Note that unlike Questions #1-3, Questions #4 and 5 were measured using a scale from 1-7 with 7 corresponding to "very effective" and 1 corresponding to "not at all effective."

a small effect. However, responses representing the perceived magnitude of the mnemonic benefits of restudying did exhibit a significant decrease from pre-test (M = 5.000, s = 1.664) to post-test (M = 3.814, s = 1.52), t (26) = 5.089, p < .001, d = .979, a large effect. Prior to the intervention, 66.67% of students reported more appreciation for the magnitude of self-testing's memory benefit than for restudying, but after the intervention, 92.59% did so, a non-significant increase, χ^2 (1) = 1.139, p > .05.These results largely support our predictions, and therefore support the hypothesis that the Quiz-Me intervention was effective at increasing appreciation of the testing effect.

Hypothesis 2: Increased Self-Testing Usage

It was hypothesized that the Quiz-Me intervention would cause students to increase use of the self-testing strategy and decrease use of the restudy strategy. Therefore, it was predicted that the students' average rank of response option B would be significantly lower – and their average rank of response option A would be significantly higher – on SSS Question #1 ("Rank the top 5 strategies in order of how often you used them") after the intervention than before it. Response patterns for restudying and self-testing strategies on Question 1 are provided in Figure 4. For this question, response option B (self-testing) exhibited a significant decrease in rank from pre-test to post-test, t (26) = 4.328, p <.001, d = .833, a large effect. Concurrently, response option A (restudying) exhibited a significant increase in rank from pre-test to post-test, t (26) = 2.681, p <.05, d = .60, a large effect. Prior to the intervention, 29.63% of students reported using (i.e., gave a lower, more

prioritized rank for) self-testing more than restudying, but after the intervention, 74.07% did so, a significant increase, χ^2 (1) = 5.142, $p < .05$. These results confirm our predictions, and support the hypothesis that the Quiz-Me intervention was effective at increasing use of the self-testing strategy while decreasing use of the restudy strategy.

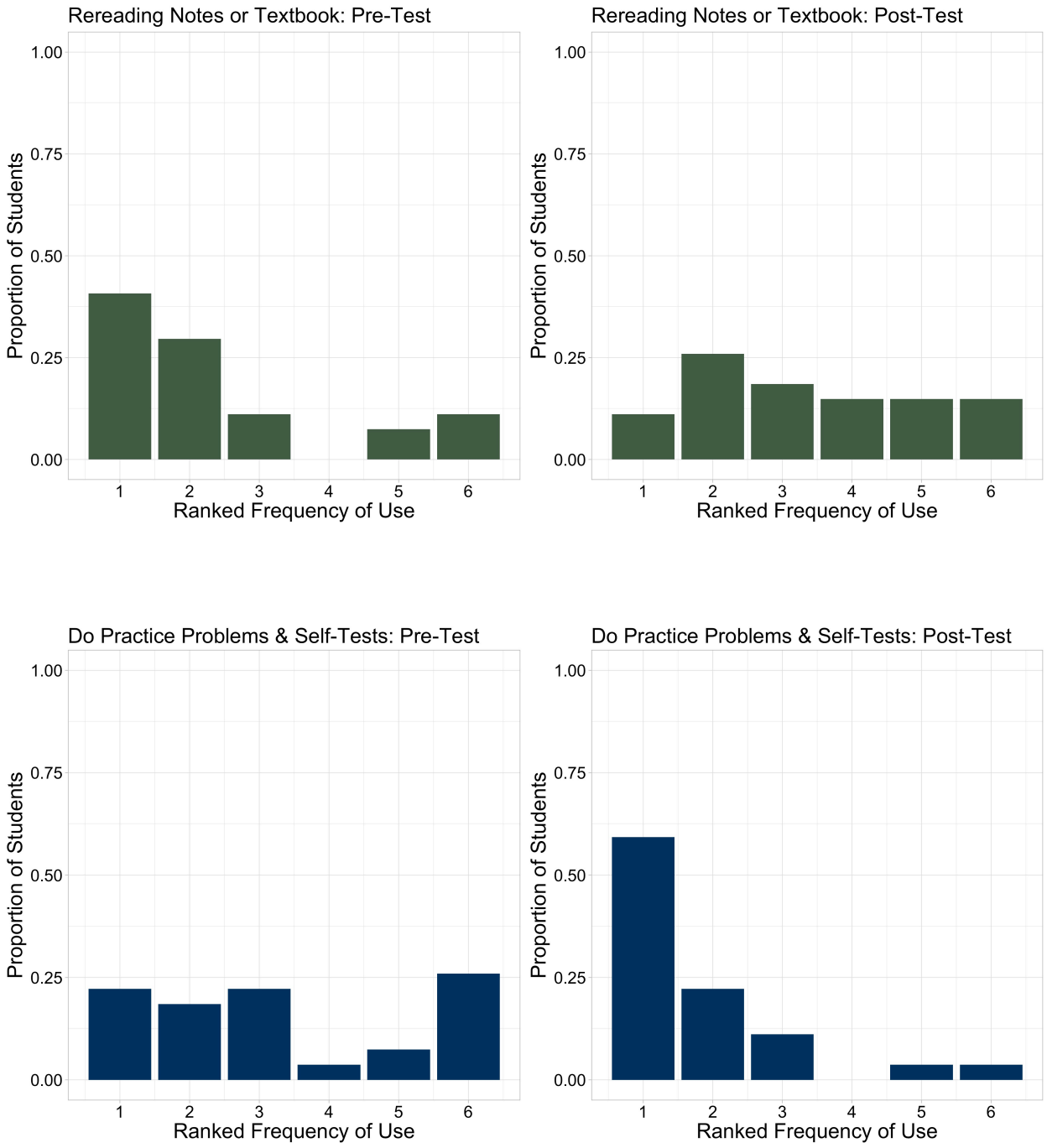


Figure 4. Study Strategy Survey responses for Experiment 1: Question #1

Hypothesis 3: Individual Differences Correlations

It was hypothesized that susceptibility to the Quiz-Me intervention was neither ubiquitous nor random, but rather that individual differences in psychological characteristics (measured in the first week of the class) were predictive of testing effect-related improvements as a consequence of experiencing the intervention. Therefore, it was predicted that some scales within the MSLQ would exhibit significant correlations with pre-to-post changes in the Study Strategy Survey questions. However, there were no specific predictions about which scales or which testing-effect related improvements would be associated with each other, so this analysis was entirely exploratory in nature. Furthermore, due to the combination of the high number of correlation tests that would be analyzed and the small sample size, inferential statistical tests are not reported, although correlation coefficients are provided and speculatively interpreted.

Correlations were tested between MSLQ scales and Pre-to-Post differences for SSS Questions #1-5. Positive difference scores for Questions #1-3 indicate increased numerical rank over time, corresponding to decreased prioritization. Conversely, decreased numerical rank over time corresponds to increased prioritization. Therefore, positive correlations between psychological characteristics and difference scores for these questions indicate that higher scores on that survey scale are associated with deprioritization of that strategy. Conversely, negative correlations between psychological characteristics and difference scores for these questions indicate that higher scores on that survey scale are associated with

increased prioritization of that strategy. A useful heuristic for interpreting results for Questions 1-3 is to treat positive correlations as negative correlations, and vice-versa. Correlation coefficients between MSLQ scales and pre-intervention SSS scores are provided in Table 3**.** Correlation coefficients between MSLQ scales and pre/post-intervention SSS difference scores are provided in Table 4**.** Correlation coefficients between pre- and post-intervention scores and pre- and post-intervention difference scores are provided in Table 5.

Question #1 A difference scores representing rank change in the use of restudying after the intervention were negatively correlated[10] with critical thinking, elaboration, and technology embrace, whereas task value and metacognition exhibited slightly weaker negative correlations. In other words, having higher values on these scales was associated with more adoption of restudying after the intervention than before it. Conversely, having lower values on these scales was associated with less adoption of restudying after the intervention than before it. No substantial (i.e., $r > .30$) positive correlations were observed. These results suggest that students low in these values were particularly likely to decrease engagement in restudying after the intervention, thereby suggesting that the intervention was especially effective for them.

[10] It is important to emphasize that these correlations should not be interpreted as statistically significant. Nevertheless, correlation coefficients with the greatest absolute value will be described in the text, as this documentation might prove useful for future research.

Table 3

Correlations between MSLQ scales and pre-intervention SSS scores in Experiment 1

	CL	CT	ER	EL	EG	HS	IG	MC	MO	OR	PL	ST	RE	SE	TV	TE	TA	TM
CL	-																	
CT	0.19	-																
ER	0.22	0.52	-															
EL	0.29	0.35	0.54	-														
EG	-0.05	-0.28	-0.38	-0.06	-													
HS	0.11	-0.04	0.03	0.09	0.09	-												
IG	0.44	0.67	0.6	0.49	-0.19	0.25	-											
MC	0.4	0.64	0.46	0.29	-0.37	-0.02	0.55	-										
MO	-0.02	0.22	0.11	-0.1	-0.13	-0.05	0.2	0.35	-									
OR	-0.02	-0.06	0.28	0.42	0.19	0.12	0.16	-0.03	0.07	-								
PL	0.3	0.25	0.17	0.21	0.36	0.7	0.44	0.12	0.06	0.3	-							
ST	0.09	0.3	0.2	0.07	-0.26	0.06	0.27	0.65	0.31	0.15	0.15	-						
RE	0.06	0.37	0.22	0.37	0.1	0.13	0.21	0.36	-0.18	0.35	0.32	0.34	-					
SE	0.35	0.31	0.59	0.31	-0.2	0.18	0.46	0.35	0.02	0.12	0.27	0.05	0.11	-				
TV	0.45	0.53	0.62	0.59	-0.09	-0.08	0.75	0.46	0.11	0.21	0.26	0.13	0.2	0.4	-			
TE	0.22	0.26	0.08	0.1	0.13	0.05	0.11	-0.01	-0.02	-0.1	0.15	-0.27	0.02	0.38	0.1	-		
TA	-0.21	0.13	-0.53	-0.23	0.33	-0.33	-0.06	0.01	0.3	-0.26	-0.15	0.03	-0.1	-0.54	-0.08	-0.01	-	
TM	0.43	0.39	0.7	0.35	-0.46	0.29	0.62	0.64	0.17	0.14	0.25	0.36	0.24	0.72	0.5	0.11	-0.6	-
T1 Q1 A	0.56	0.32	0.37	0.53	-0.25	0.06	0.59	0.36	-0.19	-0.13	0.01	0.1	0.06	0.42	0.51	0.02	-0.23	0.52
T1 Q1 B	-0.05	0.08	0.29	-0.03	0.05	-0.02	0.11	-0.15	-0.12	-0.01	-0.04	-0.52	-0.08	0.28	0.03	0.17	-0.28	0.19
T1 Q2 A	0.33	-0.09	0.2	0.36	-0.02	-0.22	0.33	0.03	-0.03	0.2	-0.09	0.03	-0.28	0.2	0.4	-0.05	0.01	0.12
T1 Q2 B	-0.06	0.28	0.31	0.3	-0.03	0.39	0.36	0.13	-0.13	-0.05	0.34	-0.26	0.13	0.25	0.27	-0.01	-0.27	0.38
T1 Q3 A	0.71	0.09	0.12	0.23	-0.14	-0.09	0.34	0.16	-0.25	0.04	0.02	-0.02	-0.17	0.32	0.31	0.01	-0.25	0.23
T1 Q3 B	0.18	-0.08	0	0.09	0.3	0.41	0.24	-0.07	-0.13	0.18	0.37	-0.18	0.22	0.17	0.09	0.15	-0.24	0.25
T1 Q4	-0.38	0.07	-0.05	-0.27	-0.14	-0.14	-0.22	0.32	0.15	-0.18	-0.25	0.3	0.23	-0.2	-0.27	-0.13	0.18	0.01
T1 Q5	-0.03	0.21	0.03	0.31	0.11	-0.1	0.06	-0.04	-0.09	-0.17	0.03	0.05	0.12	0.01	0.36	0.14	0.16	-0.14

CL = Control of Learning, CT = Critical Thinking, ER = Effort Regulation, EL = Elaboration, EG = Extrinsic Goal Orientation, HS = Help Seeking, IG = Intrinsic Goal Orientation, MC = Metacognition, MO = Motivation, OR = Organization, PL = Peer Learning, ST = Self-Testing, RE = Rehearsal, SE = Self-Efficacy, TV = Task Value, TE = Tech Embrace, TA = Test Anxiety, TM =Time Management

T1 Q1 A = Pre-test Ranked Use of Restudying, T1 Q1 B = Pre-test Ranked Use of Self-Testing, T1 Q2 A = Pre-test Ranked Appreciation of Restudying for Other, T1 Q2 B = Pre-test Ranked Appreciation of Self-Testing for Other, T1 Q3 A = Pre-test Ranked Appreciation of Restudying for Self, T1 Q3 B = Pre-test Ranked Appreciation of Self-Testing for Self, T1 Q4 = Pre-test Perceived Magnitude of Memory Benefit for Restudying, T1 Q5 = Pre-test Perceived Magnitude of Memory Benefit for Self-Testing

Note: Higher ranks for T1 Q1-Q3 represent lower priority, so positive correlations with these items indicates that higher personality characteristics are associated with lower prioritization

Table 4

Correlations between MSLQ scales and pre/post-intervention SSS difference scores in Experiment 1

	CL	CT	ER	EL	EG	HS	IG	MC	MO	OR	PL	ST	RE	SE	TV	TE	TA	TM
CL	-																	
CT	0.19	-																
ER	0.22	0.52	-															
EL	0.29	0.35	0.54	-														
EG	-0.05	-0.28	-0.38	-0.06	-													
HS	0.11	-0.04	0.03	0.09	0.09	-												
IG	0.44	0.67	0.6	0.49	-0.19	0.25	-											
MC	0.4	0.64	0.46	0.29	-0.37	-0.02	0.55	-										
MO	-0.02	0.22	0.11	-0.1	-0.13	-0.05	0.2	0.35	-									
OR	-0.02	-0.06	0.28	0.42	0.19	0.12	0.16	-0.03	0.07	-								
PL	0.3	0.25	0.17	0.21	0.36	0.7	0.44	0.12	0.06	0.3	-							
ST	0.09	0.3	0.2	0.07	-0.26	0.06	0.27	0.65	0.31	0.15	0.15	-						
RE	0.06	0.37	0.22	0.37	0.1	0.13	0.21	0.36	-0.18	0.35	0.32	0.34	-					
SE	0.35	0.31	0.59	0.31	-0.2	0.18	0.46	0.35	0.02	0.12	0.27	0.05	0.11	-				
TV	0.45	0.53	0.62	0.59	-0.09	-0.08	0.75	0.46	0.11	0.21	0.26	0.13	0.2	0.4	-			
TE	0.22	0.26	0.08	0.1	0.13	0.05	0.11	-0.01	-0.02	-0.1	0.15	-0.27	0.02	0.38	0.1	-		
TA	-0.21	0.13	-0.53	-0.23	0.33	-0.33	-0.06	0.01	0.3	-0.26	-0.15	0.03	-0.1	-0.54	-0.08	-0.01	-	
TM	0.43	0.39	0.7	0.35	-0.46	0.29	0.62	0.64	0.17	0.14	0.25	0.36	0.24	0.72	0.5	0.11	-0.6	-
Q1 A Dif	-0.29	-0.47	-0.27	-0.47	0.09	0.16	-0.33	-0.35	0	0.22	0.02	-0.11	-0.08	-0.33	-0.37	-0.42	-0.06	-0.26
Q1 B Dif	-0.06	0.03	-0.3	0	-0.12	-0.16	-0.19	0.13	0.18	-0.17	-0.22	0.33	0.01	-0.38	-0.2	-0.1	0.37	-0.28
Q2 A Dif	-0.04	0.11	-0.12	-0.45	-0.19	0.19	-0.07	-0.01	0.09	-0.26	0.06	-0.21	-0.06	-0.06	-0.14	0.06	-0.05	0.07
Q2 B Dif	-0.14	-0.29	-0.43	-0.35	-0.02	-0.28	-0.55	-0.27	0.14	0.04	-0.35	0.18	-0.04	-0.35	-0.53	-0.09	0.22	-0.41
Q3 A Dif	-0.38	-0.07	0.03	-0.21	-0.01	0.13	-0.1	-0.14	0.21	-0.1	-0.09	-0.15	-0.03	-0.2	0.05	-0.04	0.18	-0.06
Q3 B Dif	-0.19	0	0.08	-0.05	-0.38	-0.41	-0.23	0.07	0.1	-0.1	-0.43	0.09	-0.26	-0.13	-0.05	-0.2	0.12	-0.16
Q4 Dif	0.09	-0.14	0.1	0.33	0.25	0.01	0.01	-0.25	0.08	0.41	0.18	-0.17	-0.1	0	0.12	0.2	-0.18	-0.05
Q5 Dif	0.26	-0.12	0.06	-0.29	-0.05	0.07	0.08	0.15	0.12	0.11	0.08	0.07	-0.12	0.22	-0.25	0.01	-0.15	0.19

CL = Control of Learning, CT = Critical Thinking, ER = Effort Regulation, EL = Elaboration, EG = Extrinsic Goal Orientation, HS = Help Seeking, IG = Intrinsic Goal Orientation, MC = Metacognition, MO = Motivation, OR = Organization, PL = Peer Learning, ST = Self-Testing, RE = Rehearsal, SE = Self-Efficacy, TV = Task Value, TE = Tech Embrace, TA = Test Anxiety, TM =Time Management

Q1 A Dif = Change in Ranked Use of Restudying, Q1 B Dif = Change in Ranked Use of Self-Testing, Q2 A Dif = Change in Ranked Appreciation of Restudying for Other, Q2 B Dif = Change in Ranked Appreciation of Self-Testing for Other, Q3 A Dif = Change in Ranked Appreciation of Restudying for Self, Q3 B Dif = Change in Ranked Appreciation of Self-Testing for Self, Q4 Dif = Change in Perceived Magnitude of Memory Benefit for Restudying, Q5 Dif = Change in Perceived Magnitude of Memory Benefit for Self-Testing

Note: Positive scores for Q1-Q3 Difs represent increased rank and decreased priority, so positive correlations with these items indicates that higher personality characteristics are associated with decreased prioritization over time.

Table 5

Correlations between task pre-intervention, post-intervention, and SSS difference scores in Experiment 1

	T1 Q1 A	T1 Q1 B	T1 Q2 A	T1 Q2 B	T1 Q3 A	T1 Q3 B	T1 Q4	T1 Q5	T2 Q1 A	T2 Q1 B	T2 Q2 A	T2 Q2 B	T2 Q3 A	T2 Q3 B	T2 Q4	T2 Q5	Q1 A Dif	Q1 B Dif	Q2 A Dif	Q2 B Dif	Q3 A Dif	Q3 B Dif	Q4 Dif	Q5 Dif
T1 Q1 A	-																							
T1 Q1 B	-0.04	-																						
T1 Q2 A	0.36	-0.01	-																					
T1 Q2 B	0.25	0.55	-0.14	-																				
T1 Q3 A	0.56	0.04	0.57	-0.09	-																			
T1 Q3 B	0.01	0.57	-0.02	0.5	0.02	-																		
T1 Q 4	-0.16	-0.15	-0.58	-0.1	-0.57	-0.26	-																	
T1 Q 5	0.24	-0.44	-0.17	-0.14	-0.18	-0.29	0.11	-																
T2 Q1 A	0.29	0.33	0.32	0.25	0.37	0.46	-0.25	-0.45	-															
T2 Q1 B	0.09	0.39	-0.18	0.24	0	-0.17	0.13	-0.26	-0.1	-														
T2 Q2 A	0.27	0.21	0.28	0.2	0.51	-0.04	-0.4	-0.33	0.32	0.21	-													
T2 Q2 B	-0.21	0.16	-0.32	0.17	-0.26	0.08	0.08	-0.19	0.03	0.29	-0.22	-												
T2 Q3 A	0.26	0.34	0.42	0.14	0.38	0.28	-0.38	-0.15	0.62	-0.14	0.59	-0.14	-											
T2 Q3 B	0.15	0.31	0.19	0.36	-0.07	0.14	0.07	-0.36	0.45	0.06	0.24	-0.11	0.26	-										
T2 Q4	-0.19	-0.28	-0.5	-0.13	-0.5	-0.34	0.71	0.15	-0.52	0.23	-0.34	-0.08	-0.68	0.04	-									
T2 Q5	0.07	-0.21	0.09	-0.41	0.21	-0.1	-0.15	0.47	-0.34	-0.38	-0.21	-0.47	-0.13	-0.56	-0.14	-								
Q1 A Dif	-0.62	0.3	-0.05	-0.01	-0.18	0.37	-0.07	-0.57	0.58	-0.16	0.04	0.2	0.29	0.24	-0.26	-0.34	-							
Q1 B Dif	0.1	-0.77	-0.11	-0.4	-0.04	-0.71	0.25	0.28	-0.41	0.28	-0.07	0.04	-0.45	-0.29	0.45	-0.05	-0.42	-						
Q2 A Dif	-0.13	0.17	-0.69	0.27	-0.13	-0.01	0.22	-0.1	-0.04	0.33	0.5	0.13	0.07	0.01	0.2	-0.24	0.07	0.05	-					
Q2 B Dif	-0.35	-0.41	-0.06	-0.81	-0.07	-0.41	0.14	0.01	-0.21	-0.05	-0.31	0.44	-0.21	-0.4	0.07	0.1	0.13	0.39	-0.17	-				
Q3 A Dif	-0.33	0.24	-0.21	0.2	-0.66	0.21	0.24	0.05	0.15	-0.12	-0.01	0.14	0.45	0.28	-0.07	-0.31	0.4	-0.33	0.18	-0.1	-			
Q3 B Dif	0.04	-0.45	0.09	-0.36	-0.05	-0.94	0.28	0.15	-0.29	0.19	0.13	-0.12	-0.18	0.22	0.35	-0.1	-0.28	0.6	0.01	0.26	-0.1	-		
Q4 Dif	-0.02	-0.14	0.17	-0.02	0.16	-0.07	-0.48	0.04	-0.31	0.12	0.13	-0.21	-0.33	-0.04	0.27	0.03	-0.23	0.23	-0.06	-0.11	-0.42	0.05	-	
Q5 Dif	-0.22	0.36	0.25	-0.11	0.34	0.26	-0.22	-0.82	0.29	0.04	0.23	-0.08	0.09	0.04	-0.27	0.12	0.43	-0.35	-0.05	0.05	-0.26	-0.24	-0.03	-

T1 Q1 A = Pre-test Ranked Use of Restudying, T1 Q1 B = Pre-test Ranked Use of Self-Testing, T1 Q2 A = Pre-test Ranked Appreciation of Restudying for Other, T1 Q2 B = Pre-test Ranked Appreciation of Self-Testing for Other, T1 Q3 A = Pre-test Ranked Appreciation of Restudying for Self, T1 Q3 B = Pre-test Ranked Appreciation of Self-Testing for Self, T1 Q4 = Pre-test Perceived Magnitude of Memory Benefit for Restudying, T1 Q5 = Pre-test Perceived Magnitude of Memory Benefit for Self-Testing
T2 Q1 A = Post-test Ranked Use of Restudying, T2 Q1 B = Post-test Ranked Use of Self-Testing, T2 Q2 A = Post-test Ranked Appreciation of Restudying for Other, T2 Q2 B = Post-test Ranked Appreciation of Self-Testing for Other, T2 Q3 A = Post-test Ranked Appreciation of Restudying for Self, T2 Q3 B = Post-test Ranked Appreciation of Self-Testing for Self, T2 Q4 = Post-test Perceived Magnitude of Memory Benefit for Restudying, T2 Q5 = Post-test Perceived Magnitude of Memory Benefit for Self-Testing
Q1 A Dif = Change in Ranked Use of Restudying, Q1 B Dif = Change in Ranked Use of Self-Testing, Q2 A Dif = Change in Ranked Appreciation of Restudying for Other, Q2 B Dif = Change in Ranked Appreciation of Self-Testing for Other, Q3 A Dif = Change in Ranked Appreciation of Restudying for Self, Q3 B Dif = Change in Ranked Appreciation of Self-Testing for Self, Q4 Dif = Change in Perceived Magnitude of Memory Benefit for Restudying, Q5 Dif = Change in Perceived Magnitude of Memory Benefit for Self-Testing

Question #1 B difference scores representing rank change in the use of self-testing after the intervention were negatively correlated with self-efficacy. Test anxiety exhibited a slightly weaker positive correlation with Question 1 B difference scores. In other words, having higher values on the self-efficacy scale, and lower values on the test anxiety scale, was associated with more adoption of self-testing after the intervention than before it. Conversely, having lower values on self-efficacy and higher values on test anxiety was associated with less adoption of self-testing after the intervention than before it. These results suggest that students high in self-efficacy and low in test anxiety were particularly likely to increase engagement in self-testing after the intervention, thereby suggesting that the intervention was especially effective for them.

Question #2 A difference scores representing rank change in the appreciation of restudying for the average student[11] after the intervention were negatively correlated with elaboration. In other words, having higher values on elaboration was associated with more appreciation of restudying for the average student after the intervention than before it. Conversely, having lower values on elaboration was associated with less appreciation of restudying for the average student after the intervention than before it. No substantial positive correlations were observed. This result complements the interpretation of Question 1 A by suggesting that students low in elaboration were particularly likely to decrease their appreciation for

[11] Note that "for the average student" was part of the wording for Question #3 and does not describe statistical averaging across students, as might otherwise be interpreted.

restudying as a memory benefit for the average student after the intervention, thereby suggesting that the intervention was especially effective for them.

Question #2 B difference scores representing rank change in the appreciation of self-testing for the average student after the intervention were negatively correlated with intrinsic goal orientation and task value, as well as effort regulation and time management, whereas peer learning, self-efficacy, and elaboration exhibited slightly weaker negative correlations. In other words, having higher values on these scales was associated with more appreciation of self-testing for the average student after the intervention than before it. Conversely, having lower values on these scales was associated with less appreciation of self-testing for the average student after the intervention than before it. No substantial positive correlations were observed. This result suggests that students high in these values were particularly likely to increase their appreciation for self-testing as a memory benefit for the average student after the intervention, thereby suggesting that the intervention was especially effective for them.

Question #3 A difference scores representing rank change in the appreciation of restudying for oneself after the intervention were negatively correlated with control of learning. In other words, having higher values on control of learning was associated with more appreciation of restudying for oneself after the intervention than before it. Conversely, having lower values on control of learning was associated with less appreciation of restudying for oneself after the intervention than before it. No substantial positive correlations were observed. This result

suggests that students low in control of learning were particularly likely to decrease their appreciation for restudying as a memory benefit for oneself after the intervention, thereby suggesting that the intervention was especially effective for them.

Question #3 B difference scores representing rank change in the appreciation of self-testing for oneself after the intervention were negatively correlated with extrinsic goal orientation, help-seeking, and peer learning. In other words, having higher values on these scales was associated with more appreciation of self-testing for oneself after the intervention than before it. Conversely, having lower values on these scales was associated with less appreciation of self-testing for oneself after the intervention than before it. No substantial positive correlations were observed. This result suggests that students high in these values were particularly likely to increase their appreciation for self-testing as a memory benefit for oneself after the intervention, thereby suggesting that the intervention was especially effective for them.

Unlike for Questions #1-3, difference scores used in correlations for Questions #4 and 5 were measured not from ranking items but rather from questions assessing the perceived magnitude of the mnemonic benefits of the two most relevant strategies. Therefore, positive difference scores represent increased perceptions of benefits, and positive correlations with psychological characteristics indicate that higher values on that characteristics are associated with greater increases in perceived mnemonics benefits for the relevant strategy.

Question #4 difference scores representing change in the perception of mnemonic benefits for restudying after the intervention were positively correlated with organization and exhibited a slightly weaker correlation with elaboration. In other words, having higher values on these scales was associated with increased perception of mnemonic benefits for restudying after the intervention than before it. Conversely, having lower values on these scales was associated with decreased perception of mnemonic benefits for restudying after the intervention than before it. No substantial (i.e., $|r| > .30$) negative correlations were observed. This result suggests that students low in these scales were particularly likely to decrease their perception of mnemonic benefits for restudying after the intervention, thereby suggesting that the intervention was especially effective for them. Unfortunately, no substantial correlations were observed between Question #5 difference scores representing change in the perception of mnemonic benefits for self-testing after the intervention and psychological characteristics on the MSLQ.

A final exploratory analysis was conducted to determine if any insight could be gained into why subjects differed in their rankings of self-testing with regard to the best learning outcomes for the average student (Question #2 B) compared to the best learning outcomes for themselves (Question #3 B). Would any psychological characteristics predict the tendency to acknowledge the value of self-testing for other students but not for themselves, or vice-versa? Survey scales that exhibited substantial correlations with either Question #2 B or #3 B, but not both, were each entered into distinct linear models, with Pre/Post SSS-item change-by-item type

interaction entered as the dependent variable.[12] Two models that were observed to have marginally significant interaction terms are displayed in Table 6. Using estimates from these models, predictions were made regarding the distinction between appreciation gains from pre-to-post-test for the average student and for oneself. Surprisingly, students lower in task value exhibited greater gains to appreciation for self-testing's benefits for the average student than for themselves, whereas the opposite was true for students higher in task value. On average, students one standard deviation below the mean on task value would be predicted to exhibit a 3.682 greater pre/post rank decrease (i.e., priority increase) for self-testing's ranked benefits for themselves than for the average student. Students at the mean for task value would be predicted to exhibit a 1.673 greater pre/post rank decrease (i.e., priority increase) for self-testing's ranked benefits for themselves than for the average student. Students one standard deviation above the mean on task value would be predicted to exhibit a .336 greater pre/post rank *increase* (i.e., priority *decrease*) for self-testing's ranked benefits for themselves than for the average student.

[12] This measure was created by subtracting pre-from-post intervention survey measures for both SSS Q #2 and Question #3, resulting in one difference score for each question. These two difference scores were then subtracted from each other, resulting in one final difference score that represents the two-way interaction.

Table 6
Two exploratory models that attempt to explain why students differ in their appreciation of the testing effect for others compared to themselves in Experiment 1

Predictor	Estimate	Std. Error	*t*	*p*
Model 1: Task Value by Other/Self question interaction on testing effect awareness				
(Intercept)	-11.946	5.422	-2.203	0.037
Task Value Interaction	1.849	0.839	2.203	0.037
Model 2: Effort Reg. by Other/Self question interaction on testing effect awareness				
(Intercept)	-5.031	2.150	-2.34	0.027
Effort Reg. Interaction	0.917	0.387	2.37	0.025

Similar results were found for effort regulation. On average, students one standard deviation below the mean on effort regulation would be predicted to exhibit a 1.069 greater pre/post rank decrease (i.e., priority increase) for self-testing's ranked benefits for themselves than for the average student. Students at the mean for effort regulation would be predicted to exhibit a .037 greater pre/post rank decrease (i.e., priority increase) for self-testing's ranked benefits for themselves than for the average student. Students one standard deviation above the mean on effort regulation would be predicted to exhibit a .995 greater pre/post rank *increase* (i.e., priority *decrease*) for self-testing's ranked benefits for themselves than for the average student. Although these results were statistically significant, they were not predicted in advance, and so should be considered exploratory results used to provide interesting clues for the consideration of future investigations.

Discussion

Experiment 1 was conducted to test three related hypotheses. First, it was hypothesized that the Quiz-Me intervention would increase appreciation for the testing effect. To that end, it was predicted that students exposed to the Quiz-Me online educational intervention would exhibit significant improvements to indicators of testing effect appreciation. The results of Experiment 1 confirmed this prediction. These results took the form of large and reliable increases in the prioritization of self-testing from pre-to-post-test on questions assessing the appreciation of the learning benefits of self-testing for the average student and for themselves. It is intrinsic to the testing effect that not only is self-testing beneficial for memory, but that it is superior for memory than is restudying. Accordingly, we predicted decreases in the appreciation of restudying, a prediction that was strongly confirmed. Large and reliable decreases were observed in the prioritization of restudying from pre-to-post-test on questions assessing the appreciation of the learning benefits of restudying for the average student and for themselves. Furthermore, student perceived not only the ranking of restudying worse off after the intervention than before it, but perceived the magnitude of the mnemonic benefits of restudying to be weaker after the intervention. The only evidence contrary to our predictions for the testing effect appreciation hypothesis was that the perceived magnitude of the self-testing strategy was not greater after the intervention than before it. We speculate that this null effect was due to a combination of a high initial appreciation of the self-testing strategy (i.e., a ceiling

effect) combined with a modest sample size. Students did numerically improve as predicted, as evinced by the modest observed effect size. It is quite surprising how high the initial appreciation of the benefits of self-testing are, given the extensive literature documenting students' frequent ignorance of this important learning strategy. Still, we conclude that the intervention was by and large successful in increasing awareness of the testing effect as exhibited by greater appreciation of the self-testing strategy and disillusionment with the restudy strategy.

The second hypothesis tested in the present study was that was the Quiz-Me intervention would elicit increased use of the self-testing strategy. To that end, it was predicted that students exposed to the Quiz-Me online educational intervention would exhibit significant improvements to indicators of self-testing usage. The results of Experiment 1 confirmed this prediction. These results took the form of large and reliable increases in the prioritization of self-testing from pre-to-post-test on questions assessing the use of the self-testing learning strategy. Relatedly, we observed large and reliable decreases in the use of the restudying learning strategy. Large and reliable decreases were observed in the prioritization of restudying from pre-to-post-test on questions assessing the appreciation of the learning benefits of restudying for the average student and for themselves. It is worth noting that here and elsewhere, the rankings between self-testing and restudy strategy are not independent, as increased prioritization of self-testing had a direct influence upon prioritization measures of other strategies due to the nature of the ranking system used. Therefore, detriments to the restudy strategy should not be viewed as

convergent evidence in support of the relevant hypotheses but rather as the same results presented from a different angle. Nevertheless, we conclude that the intervention presents a clear success in the attempt to increase use of the self-testing strategy.

The third and final hypothesis in the present study was that the gains made in testing effect awareness and self-testing usage are related to individual differences in psychological characteristics. We predicted that some of the scales present in the MSLQ would exhibit significant correlations with the dependent variables of the previous two hypotheses. Trends were found suggesting that change in ranked appreciation of the testing effect for the average student was strongest among subjects: low in control of learning, high in extrinsic goal orientation, high in help-seeking, and high in peer-learning. The change in ranked appreciation of the testing effect for oneself was strongest among subjects: low in elaboration, high in intrinsic goal orientation, high in task value, high in effort regulation, and high in time management. The change in ranked use of the self-testing strategy was strongest among subjects: low in critical thinking, low in elaboration, low in technology embrace, low in test anxiety, and high in self-efficacy. The depreciation in perceived magnitude of the mnemonic benefits of restudying was strongest among subjects low in organization. To summarize, results suggest that the idealized subject who would be expected to experience the strongest benefits of the Quiz-Me intervention is someone who is high in: intrinsic goal orientation, extrinsic goal orientation, help-seeking, peer-learning, task value, effort regulation, time

management, and self-efficacy, and someone who is low in: control of learning, elaboration, technology embrace, test anxiety, critical thinking, and organization. These findings could be used to help select samples that are particularly susceptible to the benefits of the Quiz-Me educational intervention. Lastly, exploratory analyses revealed (and future research should attempt to confirm) that task value and effort regulation might predict the perception that benefits of self-testing are different for oneself and for other students; surprisingly, students low on these scales are likely to gain in testing effect appreciation for themselves more than for the average student. Together, these results constitute the first known attempt at understanding the individual differences behind susceptibility to the benefits of educational interventions of the testing effect.

Results of Experiment 1 clearly indicate that the intervention succeeded in raising appreciation of the testing effect and increasing use of the self-testing strategy. One of the more surprising findings from this study is the amount of incoming appreciation that students seemed to have for the testing effect and the benefits associated with self-testing prior to the intervention. As previous research has indicated (e.g., Kornell & Bjork, 2007), this level of awareness is uncommon (yet highly encouraging). One potential cause for this high level of incoming appreciation of the testing effect is that the University of Colorado Boulder psychology instructors have been known to emphasize the testing effect in introductory psychology classes. It is possible that many of the students within this sample had previous exposure to the principles of the testing effect from these sources.

Alternatively, perhaps this high level of testing effect awareness is in fact normal among students but the conversion of this appreciation into action is the substantive obstacle hindering wide-scale acceptance by students. This interpretation is given credibility by the observation that, at pre-test, students ranked self-testing as better for their learning outcomes (and for those of the average student) yet still ranked restudy as more commonly used than self-testing. This discrepancy was rectified by the Quiz-Me intervention, at which point students both appreciated the testing effect and used self-testing more than restudy.

Unfortunately, several notable limitations were present in the current study that preclude stronger conclusions from being drawn. Foremost among these considerations is the fact that this intervention occurred as an extra-credit assignment in an "Introduction to Cognitive Psychology" course, and the students were exposed to the human memory chapter between the conducting of the pre-and-post-tests. This section of the course included one lecture on how to use the principles of cognitive psychology to learn and memorize more effectively, so it is impossible to disentangle the effects of the intervention and these lectures. This lecture-based accompaniment to a testing-effect intervention was also used by Einstein et al. (2012) and by Dobson and Linderholm (2015). However, the ultimate goal of this intervention is to provide an on-demand self-testing assignment that does not take up class time, so there are clear improvements to be desired in a follow-up study. A final limitation of this study is that it, for reasons of sample size, did not include a control group, an important component of intervention

assessments that would rule out demand characteristics and allow for the evaluation of course exam grades. Experiment 2 was conducted to enact these procedural improvements.

EXPERIMENT TWO

Experiment 2 improved upon Experiment 1 in four ways. First, Experiment 2 was conducted across seven college courses in four universities in order to draw generalizable conclusions from a large and geographically diverse sample. Second, Experiment 2 used a modified Quiz-Me intervention that incorporated various functional improvements, foremost among them being the random assignment of subjects to an experimental group (as were all subjects in Experiment 1) or a control group (new to Experiment 2). Third, due to the inclusion of the control group, Experiment 2 was able to assess the effects of the Quiz-Me intervention on course exam grades. Finally, Experiment 2 was a pre-registered study within the Open Science Framework,[13] which required that all hypotheses, procedures, and analyses be specified in advanced.

Method

Participants

Seven college instructors agreed to assign the Quiz-Me intervention and the Study Strategy Survey (SSS) to their courses. Four of these instructors attend the University of Colorado Boulder, three of which are within the Psychology and Neuroscience department, and the fourth is within the Integrative Physiology department. The other three instructors attend the University of Massachusetts Dartmouth, the University of California San Diego, and Western Kentucky University, all within their respective Psychology departments. The participating

[13] See our preregistration page at https://osf.io/6mdur/

courses, their sizes of original enrollment, and their number of contributed consenting participants are listed in Table 7.

Table 7
Instructors and classes participating in Experiment 2

University	Instructor	Course	Enrollment	Participants
University of Colorado Boulder	B.K.	General Psychology	199	67 (25 control)
	D.S.M.	Statistics for Psychology	182	39 (27 control)
	J.B.	General Psychology	200	79 (30 control)
	J.H.	Statistics for Physiology	38	N/A[14]
University of Massachusetts Dartmouth	R.L.A.	Biopsychology	82	41 (20 control)
University of California San Diego	A.L.	Developmental Psychology	289	162 (81 control)
Western Kentucky University	J.R.	Educational Psychology	17	N/A[15]

[14] These data were not received in time for analysis
[15] These data were invalidated by the use of improper procedures (post-test SSS was not measured at the time of Exam #2 as required) and not included in the present analyses

Students were asked for consent to collection of website data and relevant course data once upon the Quiz-Me website. It was made clear that consent was voluntary and the instructor would remain unaware of which students consented for data collection and who participated while declining consent for data collection (yet still earned course credit). It was advised that an alternative assignment (e.g., reading and summarizing an article on the testing effect) should be offered to students who did not want to participate, but this detail was left to the discretion of each instructor. In the interest of maintaining a succinct online intervention, demographic data were not requested.

Materials

Motivated Strategies for Learning Questionnaire. Unlike Experiment 1, Experiment 2 contained a very limited role for assessing individual differences in psychological characteristics. The MSLQ was not assigned to subjects because completing this survey would have greatly increased the amount of time required to complete the Quiz-Me intervention and presumably increase the amount of subject fatigue or drop-out. It was decided that it would be preferable to obtain a large sample of subjects than to collect full MSLQ data. Nevertheless, one six-item scale from the MSLQ – task value – was retained and integrated within the Day 1 online procedure.

Study Strategy Survey. The same Study Strategy Survey was used as in Experiment 1. A multiple-choice version of this survey was created for use with

automated grading software to facilitate data entry with these larger samples. This form is shown in Appendix C.

Quiz-Me.co Web-Based Intervention. The same Quiz-Me online intervention was used as in Experiment 1, although some functional improvements had been made between experiments. These changes are described in the Procedure section below.

Procedure

As in Experiment 1, the general procedure involved assigning the SSS as a pre-test along with an in-class exam. The online participation component of the intervention was announced and made available to the class on the same date that the first exam grades were announced. It was also announced that the completion deadline for the assignment was one week prior to the second in-class exam. On that second exam, the SSS was again assigned and collected. Therefore, the procedure progressed from in-class, to online, to in-class again.

In-Class Procedure. Most of the in-class procedure was identical to that used in Experiment 1. The biggest difference was that the instructors were asked to submit not only SSS data for analysis, but also the scores for the in-class exams that accompanied the SSS.

Online Procedure. Most of the online procedure was identical to that used in Experiment 1. The biggest difference was the inclusion of a control group. After consenting to data collection and then answering six survey items assessing task value with respect to the course assigning the intervention (response options ranged

from 7, “strongly agree,” to 1, “strongly disagree”), students were randomly assigned to either an experimental condition or a control condition. Within the experimental condition, the items-to-condition (tested vs restudied) mapping was randomly counterbalanced across subjects.

Control group. The control condition consisted of a simple screen displaying one of two popular press articles on the testing effect (Belluck, 2011, 1,157 words; Freemark, 2014, 1,073 words), along with the instruction “Please read the following article in its entirety” at the top of the screen. One article was provided on each of the two days. These articles were chosen because they describe the testing effect and its benefits for students, thereby equating this condition with the experimental condition in terms of demand characteristics. After clicking the “Finish Reading” button at the bottom of the screen on Day 1, subjects were told to expect a follow-up email link that they would receive in 24 hours to return and complete the assignment in order to receive credit. After reading the other article on Day 2, subjects were told that their participation was complete and their instructor would be notified that they have earned participation credit (it was left up to the discretion of the instructor how much credit to award). In this way, the control condition was exposed to a mostly typical homework assignment, such as that which could have been provided by an instructor hoping to elicit self-testing behavior in his or her students (i.e., as a “status quo” educational intervention): to read an article of relevant text in exchange for course credit. Although it was never stated

that the subjects would be required to retain the information that they had read,[16] it might be fairly assumed that most students had the expectation that reading assignments for extra credit do not come with "no strings attached." Therefore, it can be assumed that most students had read and encoded the text rather than scrolling to the "Finish Reading" button without reading the text.

Experimental group. For the experimental group, no details on Day 1 differed from those used in Experiment 1. On Day 2, immediately after the last Final Test item had been completed, the three survey items designed to elicit goal intentions ("Being a good learner is important to me," "Receiving good grades is important to me," "Saving time by studying efficiently is important to me") were displayed along with a response scale ranging from 7 ("strongly agree") to 1 ("strongly disagree"). These questions were moved to this phase of the procedure in order to make them salient immediately prior to viewing the Feedback Page.

The Feedback page incorporated several changes, ranging from visual design to information presentation (see Appendix A). The top of the Feedback Page displayed a large text heading that read "Below is a graph of your results" along with a bar graph accompanied by the following explanations (provided to ensure that even students without the ability to understand bar graphs could appreciate the benefits of the testing effect): "The first bar shows today's test performance for facts that you Took a Practice Quiz on yesterday. The second bar shows today's test performance for facts that you Restudied yesterday." The bar graph displayed below

[16] Students in the control condition could not be tested on the information they had read, otherwise their experience might have constituted a testing effect of its own!

this text did in fact demonstrate their scores on the Final Test. Underneath this section was another large text heading that read "Below is an average student's result" with another bar graph accompanied by the following explanation: "The first bar shows how the average student did on facts that they Took a Practice Quiz on yesterday. The second bar shows how the average student did on facts that they restudied yesterday." The bar graph displayed below this text was not dynamically populated (like the bar graph for the student's own data had been) but rather was populated in advance[17] to ensure that all subjects were exposed to some form of the testing effect in the form of these "average student data," with or without seeing the testing effect in their personalized data.

Below the graphs was a paragraph of text reading: "As you can see, the average student has much better memory after taking practice tests compared to only studying. If you value being an effective learner or receiving good class grades, this strategy is for you." Below this text were "Three techniques for using practice tests for better memory: 1) Use flashcards or Quizlet.com to test yourself. Don't turn them over until you've tried to recall the answer from memory. 2) Quiz your friends and have them quiz you. Try to stump each other, then review the correct answers. 3) See if you can explain difficult concepts out loud or on paper. Compare your explanation to the one from the book or your notes." After these techniques were

[17] In Experiment 1, this graph showed a 58% to 42% accuracy advantage for tested over restudied items, taken from Rohrer et al. (2010, Figure 6). In Experiment 2, this graph showed a 63% to 49% accuracy advantage for tested over restudied items, taken from a sample of subject pool data (n = 36) who used the Quiz-Me intervention in the Fall of 2017.

listed, the Feedback Page provided several questions to the subject along with text entry boxes. The questions were: "Which strategy above can you picture yourself using?" "Why that one in particular?" "When and where will you use that strategy? Be specific! What time of day? In what location?"[18] and finally: "Please state your new strategy in If-Then format. For example, If you are in (the place and time you just identified), Then you will engage in (the strategy you just chose to use)." Underneath the final text entry box read the following text: "That concludes the Tutorial experience! Unlike many other students, you now know the benefits of practice testing, and how to use it. Thanks for participating, we hope it helps!." The page then asked: "Do you believe you will use this information to help you study in the future?" with two buttons presented below. One read "Yes-I'll use this info" and the other read "No-I won't use this info." It was hoped that selecting the affirmative response would function as a foot in the door (Freedman & Fraser, 1966) and start the subjects down the road of following their behavioral intentions.[19]

Results

A total of 388 subjects provided consent and a complete set of data. As in Experiment 1, subjects occasionally (n = 31) failed to respond appropriately to the ranking questions, such as by listing several strategies with the same ranked

---

[18] Implementation intentions have been found to be most effective when they are accompanied by tactics that increase the accessibility of the opportunity cue, such as by mental simulation and visualization of the individuals' study space and the routine they typically employ when beginning a study session (Knäuper, Roseman, Johnson, & Krantz, 2009).

[19] Technical error prevented the recording of which of these buttons the subjects had selected.

numerical priority (e.g., Restudy and Self-Testing each given the rank of 1). These instances were recoded according to the same policy enacted in Experiment 1 (see example recoding in Table 2). The final samples by instructor are reported in Table 7. Samples from two courses (instructors J. R. and J. H.) were not included in the following analyses due to substantial procedural deviations or tardiness of data provision. In total, 183 subjects assigned to the control group and 205 subjects assigned to the experimental group were analyzed. Descriptive data for the restudy and self-testing items of the SSS as well as in-class exam grades are displayed by condition in Table 8.

Unlike in Experiment 1, Quiz-Me was able to record data relevant to the Final Test within the experimental group. We predicted that the testing effect would be observed, as evinced by greater performance on previously-tested than previously-restudied items. This testing effect was observed, as previously-tested items (M = 70.98%, s = 22.54%) were recalled to a significantly greater extent than previously-restudied items (M = 57.35%, s = 27.08%), t (203) = 8.935, $p < .001$, d = .625, a medium effect. The effect of counterbalance condition on magnitude of the testing effect was assessed by a 2 (item type) x 2 (counterbalance condition) linear model with repeated measures on item type (tested, restudied). The interaction was not significant, F (1, 202) < 1, $p > .05$, $\eta^2 < .001$, indicating that the counterbalance condition did not impact the magnitude of the testing effect observed. The effect of topic chosen (e.g., world history, food and drink) on magnitude of the testing effect was assessed by a 2 (item type) x 5 (content topic) linear model with repeated

measures on item type (tested, restudied). The interaction was not significant, F (4, 199) = 1.115, $p > .05$, $\eta^2 < .021$, indicating that the content topic chosen did not impact the magnitude of the testing effect observed.

Table 8
Descriptive statistics for Study Strategy Survey by condition in Experiment 2

	Experimental Group		Control Group	
Measure	Pre	Post	Pre	Post
SSS Q #1 A	2.194 (1.515)	2.557 (1.728)	2.311 (1.648)	2.502 (1.662)
SSS Q #1 B	3.089 (1.703)	3.154 (1.879)	3.216 (1.779)	2.994 (1.808)
SSS Q #2 A	3.064 (1.760)	3.606 (1.941)	3.383 (1.782)	3.572 (1.963)
SSS Q #2 B	2.422 (1.614)	2.383 (1.678)	2.611 (1.757)	2.500 (1.795)
SSS Q #3 A	2.820 (1.711)	3.328 (1.806)	3.056 (1.786)	3.226 (1.794)
SSS Q #3 B	2.582 (1.668)	2.661 (1.806)	2.672 (1.723)	2.613 (1.794)
SSS Q #4	5.285 (1.415)	5.229 (1.386)	5.133 (1.411)	5.116 (1.403)
SSS Q #5	5.985 (1.159)	6.153 (1.070)	5.988 (1.138)	6.138 (1.071)
Exam	75.553 (11.823)	80.790 (11.294)	76.163 (12.395)	78.707 (14.716)

Mean (standard deviation)
Note: Question Part A always corresponds to Restudy, whereas Question Part B always corresponds to Self-Testing.
Note: More positive responses on SSS Q1-Q3 indicate higher numerical rank and thus lower prioritization of the corresponding strategy, whereas more positive responses on SSS Q4-Q5 indicate greater appreciation for the corresponding strategy.

Hypothesis 1: Increased Testing Effect Appreciation

Interaction with Condition. It was hypothesized that the experimental condition, characterized by direct experience with the testing effect as well as the leveraging of goal, behavioral, and implementation intentions, would elicit a greater appreciation of the testing effect than would the control condition, which was only exposed to text articles on the testing effect. It was predicted that subjects' change in attitudes towards study strategies, as measured in the SSS Questions #2-5, would demonstrate increased appreciation of self-testing and decreased appreciation of restudying. Each analysis was conducted through individual linear models with the difference in scores across tests on the relevant survey items used as the dependent measure, and group inclusion treated as the between-subject predictor variable. These analyses amounted to the test of 2 (test: pre/post) by 2 (condition: experimental, control) interactions, with repeated measures on test. To preview our results, none of the predicted interactions were found to be statistically significant.

For SSS Question #2, response option B (self-testing) did not exhibit a significant test by condition interaction, $F(1, 379) < 1$, $p > .05$, $\eta^2 < .001$. Similarly, response option A (restudying) did not exhibit a significant test by condition interaction, $F(1, 378) = 2.316$, $p > .05$, $\eta^2 = .006$. Prior to the intervention, 58.20% of experimental group students reported more appreciation (i.e., gave a lower, more prioritized rank) for self-testing's normative benefit than restudying, and after the intervention, 63.18% did so, whereas 61.11% of the control group reported more

appreciation before the intervention compared to 63.89% after it. This interaction was not significant, χ^2 (1) < 1.

For SSS Question #3, response option B (self-testing) did not exhibit a significant test by condition interaction, F (1, 379) < 1, p >.05, η^2 < .001. Similarly, response option A (restudying) did not exhibit a significant test by condition interaction, F (1, 379) = 2.447, p >.05, η^2 = .006. Prior to the intervention, 51.24% of experimental group students reported more appreciation (i.e., gave a lower, more prioritized rank) for self-testing's personal benefit than restudying, and after the intervention, 59.70% did so, whereas 54.44% of the control group reported more appreciation before the intervention compared to 59.66% after it. This interaction was not significant, χ^2 (1) < 1.

For SSS Question #4 and 5, responses representing the perceived magnitude of the mnemonic benefits of self-testing did not exhibit a significant test by condition interaction, F (1, 373) < 1, p >.05, η^2 < .001. Similarly, responses representing the perceived magnitude of the mnemonic benefits of restudying did not exhibit a significant test by condition interaction, F (1, 372) < 1, p >.05, η^2 < .001. Prior to the intervention, 53.50% of experimental group students reported more appreciation (i.e., gave a lower, more prioritized rank) for the magnitude of self-testing's memory benefit than restudying, and after the intervention, 60.20% did so, whereas 53.88% of the control group reported more appreciation for the magnitude of self-testing's memory benefit before the intervention compared to 58.88% after it. This interaction was not significant, χ^2 (1) < 1.

Despite the alignment of the observed numerical trends with our predictions, results of these statistical tests refute our prediction that the pre/post differences in SSS Questions 2-5 would depend on condition. Therefore, our hypothesis - that the direct experience of self-testing and the involvement of the Model of Action Phases would lead to greater appreciation of the testing effect than reading a text article about the effect – receives no support in the form of statistically significant interactions. However, an interesting pattern was observed in the data, demonstrated in Table 9. Specifically, only the experimental group was associated with significant pre/post changes to testing effect appreciation. The experimental group exhibited a significant rank increase (i.e., a priority decrease) in appreciation of the restudy strategy, both for the average student and for oneself, and a significant increase in the perceived magnitude of learning benefits for self-testing. No similar results were found for the control group. Although not as strong as finding a statistically significant interaction, this dissociation of results between groups constitutes evidence that the experimental group was particularly effective at changing testing effect appreciation, whereas the same is not true of the control group.

Table 9.
Study Strategy Survey responses and in-class exam grades by intervention condition in Experiment 2.

SSS Question #	Strategy	Experimental		Control	
		Pre	Post	Pre	Post
		t-test of pre/post difference scores		*t*-test of pre/post difference scores	
1: Strategy Use	Restudy	2.194	2.557	2.311	2.502
		t (200) = 2.816**		*t* (179) = 1.276	
	Self-Testing	3.089	3.154	3.216	2.994
		t (200) < 1		*t* (179) = 1.331	
2: Appreciation for Average Student	Restudy	3.064	3.606	3.383	3.572
		t (200) = 3.738***		*t* (178) = 1.163	
	Self-Testing	2.422	2.383	2.611	2.500
		t (200) < 1		*t* (178) < 1	
3: Appreciation for Oneself	Restudy	2.820	3.328	3.055	3.226
		t (200) = 3.461***		*t* (179) = 1.178	
	Self-Testing	2.582	2.661	2.672	2.613
		t (200) < 1		*t* (179) < 1	
4 & 5: Magnitude of Learning Benefits	Restudy	5.285	5.229	5.133	5.116
		t (194) < 1		*t* (178) < 1	
	Self-Testing	5.985	6.153	5.988	6.138
		t (195) = 1.969*		*t* (178) = 1.633	
In-Class Exam Grades		75.553	80.790	76.163	78.707
		t (195) = 5.772***		*t* (180) = 2.515*	

Note: *p <.05, **p <.01, ***p <.001

Pre-Post Differences. As the test (pre, post) by condition (experimental, control) interactions were not significantly different from each other for any of the relevant dependent measures, we report pre-post differences collapsed across groups. As in Experiment 1, it was hypothesized that the Quiz-Me intervention would cause students to gain appreciation for the testing effect. Therefore, it was predicted that the students' average rank of response option B ("Do practice

problems, self-tests, or flashcards") would be significantly lower (i.e., more prioritized) after the intervention than before it. Similarly, it was predicted that the average rank of response option A ("Rereading notes or textbook") would be significantly higher (i.e., less prioritized) after the intervention than before it. These predictions were made for both Study Strategy Survey Question #2 ("Which do you think lead to the best learning outcomes for the average student?") and Question #3 ("Which do you think lead to the best learning outcomes for you?"). It also predicted that SSS Question #5 ("How beneficial do you believe [practice testing] is for your memory?") would be significantly higher after the intervention than before it, and that SSS Question #4 ("How beneficial do you believe [restudying] is for your memory?") would be significantly lower after the intervention than before it. Response patterns for restudying and self-testing strategies on Question 2 and 3 are provided in Figures 5 and 6, respectively.

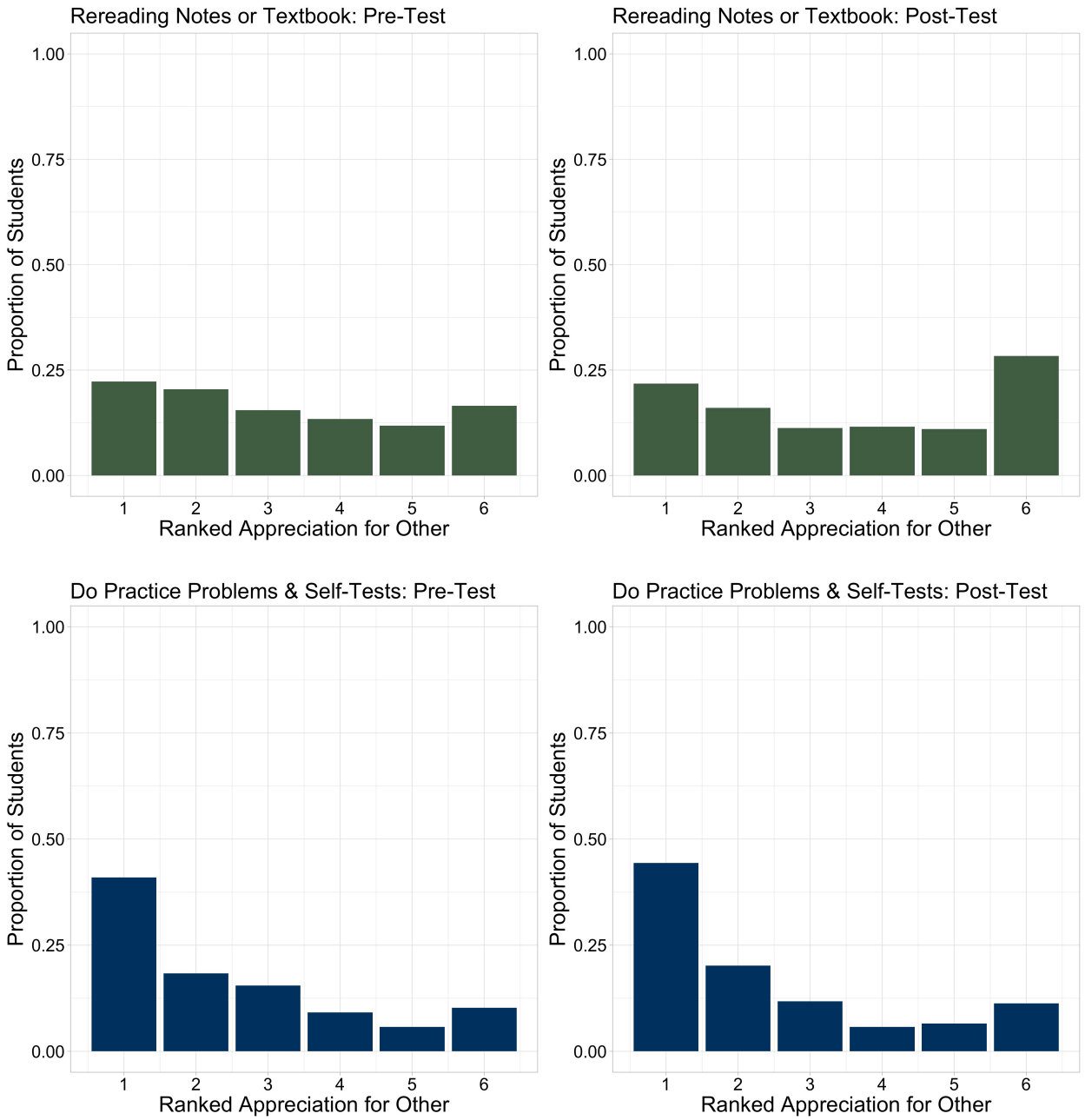


Figure 5. Study Strategy Survey responses for Experiment 2: Question #2

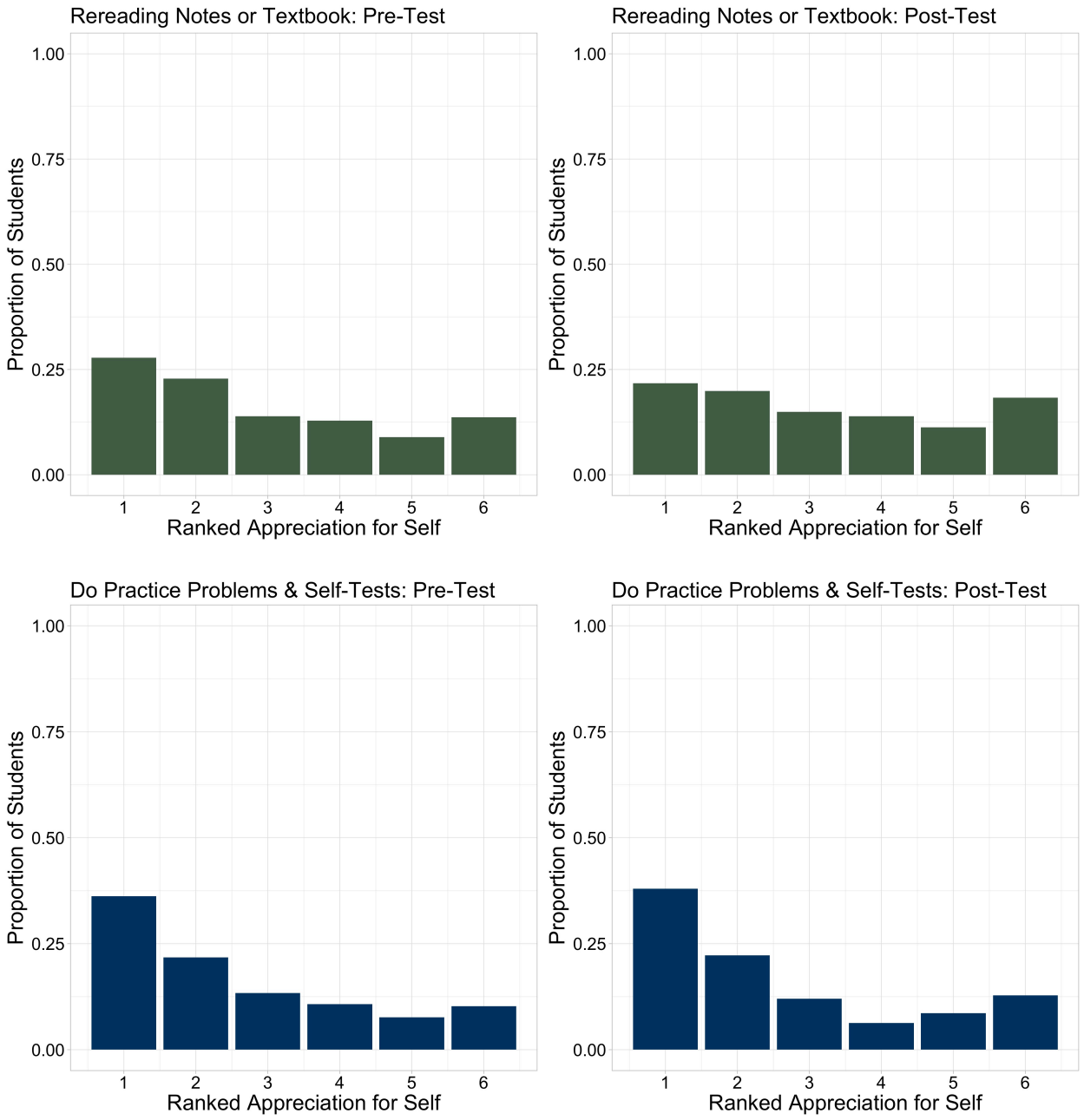


Figure 6. Study Strategy Survey responses for Experiment 2: Question #3

For SSS Question #2, response option B (self-testing) did not exhibit a significant decrease in rank from pre-test to post-test, t (379) < 1, p >.05, d = .035. However, response option A (restudying) did exhibit a significant increase in rank (i.e., a decreased in prioritization) from pre-test to post-test, t (379) = 3.404, p <.001, d = .174, a small effect. As Table 9 demonstrates, this pre/post effect for restudying was primarily driven by the experimental intervention condition. Prior to the intervention, 59.58% of students reported more appreciation for self-testing's normative benefit than restudying, and after the intervention, 63.51% did so, a non-significant increase, χ^2 (1) < 1.

For SSS Question #3, response option B (self-testing) did not exhibit a significant decrease in rank from pre-test to post-test, t (380) < 1, p > .05, d = .008. However, response option A (restudying) exhibited a significant increase in rank from pre-test to post-test, t (380) = 3.336, p <.001, d = .170, a medium effect. As Table 9 demonstrates, this pre/post effect for restudying was primarily driven by the experimental intervention condition. Prior to the intervention, 52.75% of students reported more appreciation for self-testing's personal benefit than restudying, but after the intervention, 59.68% did so, a non-significant increase, χ^2 (1) = 1.699, p > .05.

For SSS Question #4 and 5, responses[20] representing the perceived magnitude of the mnemonic benefits of self-testing did exhibit a significant increase

[20] Note that unlike Questions #1-3, Questions #4 and 5 were measured using a scale from 1-7 with 7 corresponding to "very effective" and 1 corresponding to "not at all effective."

from pre-test (M = 5.987, s = 1.148) to post-test (M = 6.151, s = 1.066), t (374) = 2.555, $p < .05$, d = .131, a small effect. As Table 9 demonstrates, this pre/post effect for self-testing was primarily driven by the experimental intervention condition. However, responses representing the perceived magnitude of the mnemonic benefits of restudying did not exhibit a significant decrease from pre-test (M = 5.221, s = 1.410) to post-test (M = 5.185, s = 5.151), t (373) <1, $p > .05$, d = .042. Prior to the intervention, 53.68% of students reported more appreciation for the magnitude of self-testing's memory benefit than restudying, but after the intervention, 59.57% did so, a non-significant increase, χ^2 (1) <1.

These results provide mixed support for our predictions. For Questions #2 and 3, results exhibited the predicted detriment to restudy priority following the intervention, but results did not exhibit the predicted improvement to self-testing. For Questions #4 and 5, results exhibited the predicted benefit of the perceived magnitude of self-testing following the intervention, but did not exhibit the predicted detriment to restudy. These findings therefore provide mixed support for the hypothesis that the Quiz-Me intervention was effective at increasing appreciation of the testing effect.

Hypothesis 2: Increased Self-Testing Usage

Interaction with Condition. It was hypothesized that the experimental condition, characterized by direct experience with the testing effect as well as the leveraging of goal, behavioral, and implementation intentions, would elicit greater use of the self-testing strategy than would the control condition, which was only

exposed to text articles on the testing effect. It was predicted that subjects' change in study behavior, as measured on the SSS Question #1, would demonstrate increased use of self-testing and decreased use of restudying. This analysis was conducted through a linear model with the difference in scores across tests on SS Question #1 responses used as the dependent measure, and group inclusion treated as the between-subject predictor variable. This analysis amounted to the test of a 2 (test: pre/post) by 2 (condition: experimental, control) interaction, with repeated measures on test. For this question, response option B (self-testing) did not exhibit a significant test by condition interaction, F (1, 379) = 1.757, p >.05, η^2 = .004. Similarly, response option A (restudying) did not exhibit a significant test by condition interaction, F (1, 379) < 1, p >.05, η^2 = .001. Prior to the intervention, 31.34% of experimental group students reported more use for self-testing than restudying, and after the intervention, 41.29% did so, whereas 34.44% of the control group reported more appreciation before the intervention compared to 40.33% after it. This interaction was not significant, χ^2 (1) <1. Despite the alignment of the observed numerical trends with our predictions, results of these statistical tests refute our prediction that the pre/post differences in SSS Questions 1 would depend on condition. Therefore, our hypothesis - that the direct experience of self-testing and the involvement of the Model of Action Phases would lead to greater use of the self-testing strategy than reading a text article about the effect – receives no support in the form of a statistically significant interaction. However, as was the case for changes in testing effect awareness, an interesting pattern was observed in

the strategy use data, demonstrated in Table 9. Specifically, only the experimental group exhibited a significant rank increase (i.e., a priority decrease) in use of the restudy strategy. No similar result was found for the control group. Although not as strong as a finding a statistically significant interaction, this dissociation of results between groups constitutes evidence that the experimental group was particularly effective at changing study strategy use, whereas the same is not true of the control group.

Pre-Post Differences. As the test (pre, post) by condition (experimental, control) interaction was not significantly different from each other, we report pre-post differences collapsed across groups. As in Experiment 1, it was hypothesized that the Quiz-Me intervention would cause students to increase use of the self-testing strategy and decrease use of the restudy strategy. Therefore, it was predicted that the students' average rank of response option B would be significantly lower – and their average rank of response option A would be significantly higher – on SSS Question #1 ("Rank the top 5 strategies in order of how often you used them") after the intervention than before it. Response patterns for restudying and self-testing strategies on Question 1 are provided in Figure 7. For this question, response option B did not exhibit a significant decrease in rank from pre-test to post-test, t (380) < 1, p > .05, d = .023. However, response option A exhibited a significant increase in rank from pre-test to post-test, t (380) = 2.862, p <.01, d = .146, a small effect. As Table 9 demonstrates, this pre/post effect for restudying was primarily driven by the experimental intervention condition. Prior

to the intervention, 32.80% of students reported using (i.e., gave a lower, more prioritized rank for) self-testing more than restudying, but after the intervention, 40.83% did so, a marginally significant increase, χ^2 (1) = 3.419, p = .069. These results provide some confirmation for our predictions, supporting the hypothesis that the Quiz-Me intervention was effective at decreasing use of the restudy strategy, despite not increasing use of the self-testing strategy.

Hypothesis 3: Increased Exam Grades

Unlike in Experiment 1, it was hypothesized that the predicted benefits of the Quiz-Me intervention would cause an increase in students' abilities to retain class-related information. Therefore, it was predicted that students would exhibit a significant pre/post-test improvement on course exam grades. Unfortunately, a test of the pre/post differences in exam grades is confounded with exam content; because instructors assessed different information on Exams #1 and 2, any main effect of pre/post-test difference could be attributed to content rather than the learners' study strategies used leading up to them.

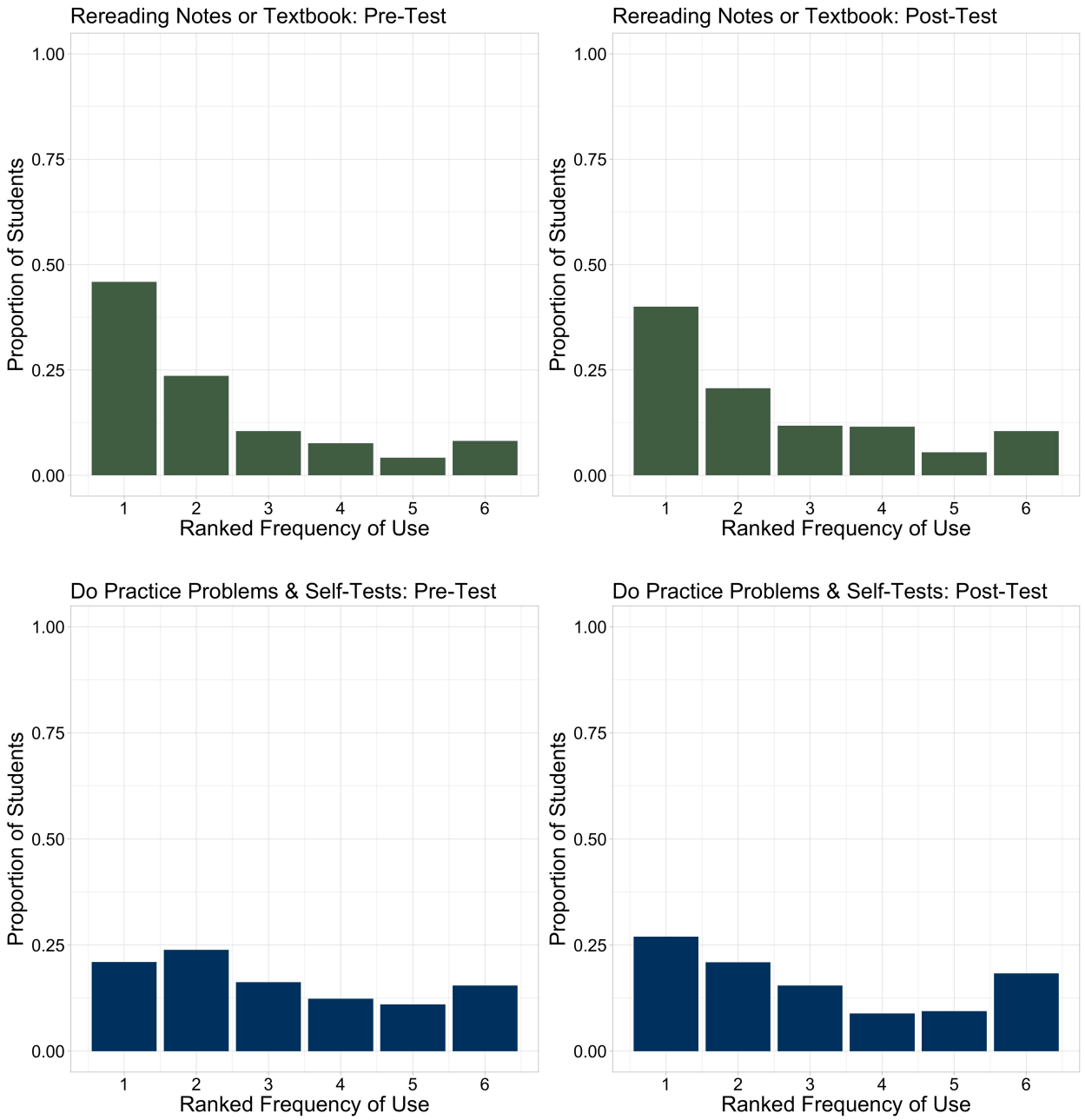


Figure 7. Study Strategy Survey responses for Experiment 2: Question #1

Interaction with Condition. To counteract this issue, a test of the exam number-by-condition interaction was analyzed. As both conditions experienced the same two exams, the test of the interaction would reveal whether either of the Quiz-Me intervention conditions made more improvements from Exam #1 to Exam #2 than the other condition. It was hypothesized that the experimental condition would elicit greater exam-preparatory improvements than would the control condition, and so it was predicted that the experimental group would exhibit greater improvement from Exam #1 to Exam #2 than would the control group. This analysis was conducted through a linear model with the difference in scores across exams used as the dependent measure, and group inclusion treated as the between-subject predictor variable. This analysis amounted to the test of a 2 (exam: first, second) by 2 (condition: experimental, control) interaction, with repeated measures on test. Results exhibited a marginally significant interaction between condition and exam differences, F (1, 371) = 2.901, p = .089, η^2 = .007. These results provide weak evidence that the effect of experimental condition had a reliable impact on the improvement in grades from Exam #1 to Exam #2. Descriptive statistics and within-condition pre/post t-tests are provided in Table 9.

Two speculative analyses were conducted to determine whether other relevant measured variables could help explain the preset data. First, the difference in instructor had an impact on the presence of the predicted condition by exam number interaction. This analysis took the form a 5 (instructor) by 2 (condition) by 2 (exam) interaction with repeated measures on exam. This interaction term was

significant, F (4, 363) = 2.584, $p < .05$, $\eta^2 = .027$, a small effect. The differences in exam improvement by instructor and condition are presented in Table 10. Notably, the only instructors for whom an advantage of experimental condition was not present was for both of the instructors teaching General Psychology classes. When removing these classes from the analysis, the condition-by-exam interaction is significant, F (1, 237) = 10.300, $p < .01$, $\eta^2 = .041$, a small effect. Importantly, this analysis of the impact of the two conditions on exam improvement was dependent on whether the course was a General Psychology course or a higher level course was motivated not by an a priori hypothesis but by ad-hoc speculation.

Table 10
The differences in exam improvement by instructor and condition in Experiment 2

Instructor	Experimental	Control	*p* (Int)	Task Value
B.K. (*n* = 67)	3.675 (8.724)	7.416 (9.554)	.120	5.553 (1.000)
D.S.M. (*n* = 39)	-8.467 (16.311)	-14.098 (20.020)	.431	5.070 (1.149)
J.B. (*n* = 79)	2.772 (13.978)	7.862 (9.782)	.092	6.018 (0.909)
R.L.A. (*n* = 41)	4.000 (9.978)	-.400 (13.320)	.222	5.705 (1.055)
A.L. (*n* = 162)	8.382 (9.672)	5.592 (9.281)	.062	5.422 (1.116)

Mean (standard deviation)
"Int" represents the exam-by-condition interaction for that instructor

It could be reasoned that the (elective) General Psychology courses could be characterized by lower task value than the higher level, psychology majors-only courses. Task value was assessed in this experiment due to its promising role in predicting positive outcomes from the intervention in Experiment 1. It was the only scale from the MSLQ that was retained from Experiment 1, and was completed by all subjects including those from the experimental and control conditions. Task value was found to differ by instructor, F (4, 365) = 6.322, p <.001, η^2 = .064, a small effect, but not in a way that would have made the latter two analyses redundant. Specifically, the predicted pattern whereby the General Psychology courses exhibited lower task value than the other courses was not observed in the data. In fact, the General Psychology courses were characterized by task value that was at or above the average of M = 5.555. We conducted an additional exploratory analysis consisting of a test of the statistical interaction between task value and condition interaction on exam grade with repeated measures on exam number. A significant task value by condition interaction was found, F (1, 353) = 4.037, p <.05, η^2 = .011, a small effect, indicating that the effect of intervention condition on exam grade improvement depends upon the level of task value exhibited by the subject. Table 11 provides estimates of model parameters and Table 12 provides predictions of grade improvement based on condition and task value. The predictions depict a pattern whereby subjects in the experimental group experience substantial test-to-test improvement regardless of task value score, whereas test-to-test improvement for control group subjects is highly dependent upon task value score. These

analyses, although entirely ad hoc in nature, support the findings of Experiment 1 that task value is an important characteristic for predicting susceptibility to the Quiz-Me intervention. More importantly, they demonstrate that subjects who are low in task value benefit by taking the experimental over the control intervention. Correlations between task value and Study Strategy Survey responses are provided in Table 13.

Table 11
Estimates of model parameters for condition by task value interaction on exam improvement in Experiment 2

Predictor	Estimate	Std. Error	*t*
Intercept	-10.280	4.787	2.147*
Condition	16.506	7.080	2.331*
Task Value	2.294	0.857	2.675**
Interaction	-2.509	1.248	2.009*

Table 12
Predictions of exam improvement based on condition by task value interaction model in Experiment 2

	Experimental	Control
Low Task Value (*M* - *s*)	5.266	-0.025
Medium Task Value (*M*)	5.033	2.467
High Task Value (*M* + *s*)	4.799	4.961

Table 13

Correlations between task value, pre-intervention, and SSS difference scores in Experiment 2

	TV	T1 Q1 A	T1 Q1 B	T1 Q2 A	T1 Q2 B	T1 Q3 A	T1 Q3 B	T1 Q4	T1 Q5	Q1 A Dif	Q1 B Dif	Q2 A Dif	Q2 B Dif	Q3 A Dif	Q3 B Dif	Q4 Dif
TV	-															
T1 Q1 A	-0.14	-														
T1 Q1 B	0.2	-0.09	-													
T1 Q2 A	-0.09	0.36	0.03	-												
T1 Q2 B	0.15	0.03	0.39	-0.18	-											
T1 Q3 A	-0.08	0.55	0	0.57	-0.09	-										
T1 Q3 B	0.18	-0.02	0.61	-0.08	0.57	-0.05	-									
T1 Q4	0.23	-0.36	0.16	-0.34	0.21	-0.47	0.18	-								
T1 Q5	0.01	0.04	-0.34	0.04	-0.31	0.07	-0.45	0.08	-							
Q1 A Dif	0.13	-0.55	0.02	-0.15	-0.01	-0.27	0.01	0.12	-0.01	-						
Q1 B Dif	-0.07	-0.01	-0.53	-0.01	-0.19	-0.02	-0.25	-0.02	0.07	-0.04	-					
Q2 A Dif	0.08	-0.18	0.04	-0.53	0.03	-0.27	0.07	0.08	-0.02	0.33	-0.02	-				
Q2 B Dif	-0.1	0.05	-0.16	0.18	-0.6	0.04	-0.25	-0.04	0.12	0	0.39	-0.14	-			
Q3 A Dif	0.12	-0.27	-0.06	-0.28	0	-0.56	0	0.15	0.02	0.51	0	0.52	-0.02	-		
Q3 B Dif	-0.12	0	-0.27	0.04	-0.28	-0.03	-0.57	-0.02	0.2	-0.06	0.5	-0.11	0.46	-0.08	-	
Q4 Dif	-0.04	0.09	-0.01	0.08	-0.07	0.17	-0.04	-0.5	-0.07	-0.18	0.11	-0.24	0.07	-0.3	0.16	-
Q5 Dif	0.06	0.05	0.1	-0.03	0.13	-0.03	0.22	-0.06	-0.58	0.04	-0.21	0.04	-0.14	0.07	-0.33	0.07

T1 Q1 A = Pre-test Ranked Use of Restudying, T1 Q1 B = Pre-test Ranked Use of Self-Testing, T1 Q2 A = Pre-test Ranked Appreciation of Restudying for Other, T1 Q2 B = Pre-test Ranked Appreciation of Self-Testing for Other, T1 Q3 A = Pre-test Ranked Appreciation of Restudying for Self, T1 Q3 B = Pre-test Ranked Appreciation of Self-Testing for Self, T1 Q4 = Pre-test Perceived Magnitude of Memory Benefit for Restudying, T1 Q5 = Pre-test Perceived Magnitude of Memory Benefit for Self-Testing
T2 Q1 A = Post-test Ranked Use of Restudying, T2 Q1 B = Post-test Ranked Use of Self-Testing, T2 Q2 A = Post-test Ranked Appreciation of Restudying for Other, T2 Q2 B = Post-test Ranked Appreciation of Self-Testing for Other, T2 Q3 A = Post-test Ranked Appreciation of Restudying for Self, T2 Q3 B = Post-test Ranked Appreciation of Self-Testing for Self, T2 Q4 = Post-test Perceived Magnitude of Memory Benefit for Restudying, T2 Q5 = Post-test Perceived Magnitude of Memory Benefit for Self-Testing
Q1 A Dif = Change in Ranked Use of Restudying, Q1 B Dif = Change in Ranked Use of Self-Testing, Q2 A Dif = Change in Ranked Appreciation of Restudying for Other, Q2 B Dif = Change in Ranked Appreciation of Self-Testing for Other, Q3 A Dif = Change in Ranked Appreciation of Restudying for Self, Q3 B Dif = Change in Ranked Appreciation of Self-Testing for Self, Q4 Dif = Change in Perceived Magnitude of Memory Benefit for Restudying, Q5 Dif = Change in Perceived Magnitude of Memory Benefit for Self-Testing

Discussion

Experiment 2 was conducted to test three related hypotheses. First, it was hypothesized that the Quiz-Me intervention would increase appreciation for the testing effect relative to before the intervention, and relative to the control intervention. To that end, it was predicted that test (pre, post) by intervention condition (experimental, control) would interact on the various measures of testing effect appreciation. This prediction was partially refuted by the finding of non-significant interactions upon testing effect appreciation, both for oneself and for the average student. Despite the lack of statistically significant interactions, pre/post differences on several measures were observed that were driven primarily by the experimental group. In fact, no pre/post differences were observed within the control group, but the experimental group repeatedly showed pre/post differences for testing effect awareness, such as reduced prioritization of the restudy strategy. Therefore the results of Experiment 2 indicate that the experimental intervention was largely effective in increasing appreciation of the testing effect, whereas the control condition was not.

Second, it was hypothesized that the Quiz-Me intervention would increase use of the self-testing strategy and decrease use of the restudy strategy, both relative to before the intervention, and relative to the control intervention. To that end, it was predicted that pre-to-post-interventions changes in restudy and self-testing use would interact with intervention group, such that experiencing the experimental group's intervention would cause larger pre-to-post changes in restudy

self-testing use than would experiencing the control group's intervention. This prediction was partially refuted by the finding of a non-significant interaction, such that pre-to-post intervention changes in restudy and self-testing use did not depend on which intervention subjects had experienced. Despite the lack of a statistically significant interaction, pre/post differences on restudy usage were observed and driven primarily by the experimental group. Specifically, no pre/post differences for restudying or self-testing were observed for the control group, but restudy use rank had significantly decreased for the experimental group. Therefore, the results of Experiment 2 indicate that the intervention was somewhat effective in changing study strategy use, namely by decreasing students' proclivity of restudying relative to other strategies, and this benefit was only experienced by the experimental intervention condition.

Third, it was hypothesized that the Quiz-Me intervention would increase students' abilities to succeed on in-class exams. However, a simple pre-to-post intervention comparison is confounded with exam content, so the only relevant prediction concerned the interaction between exam number and intervention group. These results revealed a marginally significant interaction, indicating modest support for the prediction that pre-to-post exam grade improvement depends on whether students experienced the experimental intervention or the control intervention. This benefit to exam performance might be due to the changes to study strategy and appreciation and use that was present in the experimental group but not the control group.

Additionally, ad-hoc investigation into the distribution of exam grades between instructors revealed that only the samples associated with General Psychology courses failed to display the predicted condition by exam number interaction. Additional analyses indicated that task value also exhibited a significant interaction with condition on exam improvement, indicating that the amount to which a student cares about the course at hand influences whether the experimental group will benefit him or not; when task value is high, there is little difference in outcomes by group, but when task value is low, this difference is substantial. These ad hoc analyses are not conclusive but rather might highlight important avenues for future research.

There are two primary conclusions to be drawn from Experiment 2. First, as in Experiment 1, students were receptive to taking online educational interventions and applying that information to inform their beliefs about the effectiveness of various study strategies as well as changing their study behaviors. Unlike Experiment 1, this receptivity manifested in decreased appreciation and use of restudying more so than in an increased appreciation and use of self-testing. Another primary conclusion from Experiment 2 is that students demonstrate superior intervention outcomes for the Quiz-Me experimental educational intervention than for the control intervention. Although not revealed by a significant interaction, the repeated finding that pre/post differences in restudy appreciation and use were only observed for the experimental group, combined with the marginally greater pre/post exam performance exhibited by the experimental

group, converge on the conclusion that the experimental group was successful in changing student study strategies for the better.

General Discussion

The current investigation was conducted in order to build upon the successes of previous applied research endeavors into the development and validation of educational interventions that teach students about the testing effect and elicit greater self-testing behavior. Previous studies, such as those conducted by Einstein et al. (2012) and Dobson and Linderholm (2015), have been characterized by successful educational outcomes qualified by issues of interpretation. The present experiment sought to build upon this previous research in several ways. First, previous research drew conclusions about the efficacy of interventions from either in-class exam data or retrospective surveys alone, neither of which permit the most direct assessment of attitudes and behaviors related to self-testing in isolation. The current investigation measured both in-class exam data and surveys undertaken immediately after in-class exams in order to accurately assess student study strategies and their effects on exam grades while minimizing biases of retrospection. Other important new measures included the Motivated Strategies for Learning Questionnaire, used to identified individual differences in psychological characteristics that might predict susceptibility to a successful reaction to the educational intervention. Second, in addition to the use of an experimental group defined by the presence of comparative experience and explicit recommendation of self-testing, Experiment 2 also included a control group defined by exposure to a text-based intervention in order to minimize the impact of demand characteristics and artifacts inherent to quasi-experimental procedure designs. Third, the current

investigation attempted to supplement previous intervention procedures with details designed to maximize the likelihood of behavioral change as recommended by the Model of Action Phases (Bamberg, 2013; Heckhausen & Gollwitzer, 1987), including the formation of goal, behavioral, and implementation intentions. Finally, this investigation conducted the important work of making the educational intervention an online experience, thereby preserving class time and making this testing effect elicitation experience widely accessible; its validation as an enhancement of educational outcomes would thus be an important contribution in the pursuit of effective tools to improve student study behaviors.

In many ways, the present investigation has confirmed what previous studies have found, thereby solidifying the state of the research into the elicitation of self-testing behaviors. As was previously found by Einstein et al. (2012), students in Experiment 1 confirmed a greater tendency of using self-testing behavior after the intervention than before it. As previously noted, these present data were recorded immediately after each exam rather than months in retrospect. Unlike Einstein et al., Experiment 1 found that students not only used self-testing more than before the intervention, but they ranked self-testing as better for their learning outcomes, ranked self-testing better for the average student's learning outcomes, and judged self-testing to be better in magnitude of learning outcomes than restudying. The number of students who appreciated and used self-testing more than restudying grew a substantial amount after the intervention as well. The results of Experiment 1 provide the strongest available support for the claim that educational

interventions into the testing effect can increase testing effect appreciation and self-testing use.

The results of Experiment 2 also confirm and extend the work of Einstein et al. (2012), although they do so in a more modest way than did Experiment 1. Whereas the self-testing improvements (i.e., in appreciation and use) in Experiment 1 were very large, such improvements in Experiment 2 were considerably smaller. The most plausible interpretation for these divergent results is that Experiment 1 was conducted in a course that was particularly well-suited for the intervention. As a course in cognitive psychology that covered the dynamics of learning and memory – including the testing effect – in between Exams 1 and 2, this course supplemented the direct experience of the testing effect in the online intervention with in-person lectures that emphasized many topics relevant to changing one's study strategies. This procedure is entirely in keeping with previous research, as both Einstein et al. (2012) and Dobson and Linderholm (2015) dedicated class time to discussing the testing effect and its incorporation into exam preparation routines, but this procedure is far from the fully online implementation that was used in Experiment 2. The lack of lecture-based accompaniment of the intervention may thus be the most important difference in the divergent set of results between experiments. If true, the difference in results between Experiments 1 and 2 can be attributed to the in-person lecture accompanying the intervention, and the effects present in Experiment 2 (with no in-person lecture) can be ascribed to the effects of the online procedure alone. No previous research has examined the effects of a purely online

educational intervention into the teaching of the testing effect. After only two 30-min (or shorter) sessions in the experimental condition, students exhibited significant reductions in their ranked use and appreciation of restudying, and a significant increase in the perception of the magnitude of self-testing's memory benefits. This experience can thus be viewed as an easy method for changing students' self-testing attitudes and behaviors without need for an in-person lecture component. Future research should attempt to replace not only the comparative experience component of the intervention that was conducted in person by Einstein et al. (2012) and Dobson and Linderholm (2015) and conducted here on Quiz-Me, but also to replace the in-person lectures crucial to the effects of Experiment 1 with further on-demand content such as video lecture in order to complete the transition from in-person to online intervention. In doing so, this future intervention could realize the goal of the fully on-demand testing effect intervention that Quiz-Me had attempted to achieve.

This investigation did more than confirm and extend previous research, but also drew several new and interesting conclusions. Foremost among these additions was the assessment of a control group that experienced an intervention defined not by comparative experience and goal, behavioral, and implementation intentions, but rather by reading two articles on the educational applications of the testing effect. This control group was selected in order to mimic the lay-intervention of high-school and undergraduate college teachers: assigning a passage of text and hoping that the students care enough to put the content to use. As anticipated, this control group

was less effective for changing testing effect awareness and use than the experimental condition. This change was not exhibited in significant interactions, but rather in the recurring finding that the experimental group alone was accompanied by changes to appreciation and use of the restudy strategy. Although it would have been preferable to observe changes in the use in self-testing, it is promising that Quiz-Me was successful in correcting misguided beliefs about the benefits of restudying, and in increasing the magnitude of perceived benefits of self-testing. These changes to restudying beliefs and use priorities may have caused the marginally greater exam performance that was experienced by the experimental group than the control group. Interestingly, it was found that the control group was not associated with as much exam grade improvement as the experimental group within General Psychology courses and within students low in task value. It would be valuable to determine whether these results hold in future samples as well.

The current investigation was accompanied by several notable limitations. First, although Experiment 1 identified the psychological characteristics of subjects who would be particularly susceptible to positive reactions to the educational intervention, it would be most appropriate to describe these characteristics as relevant to the Experiment 1 procedure rather than the more portable Experiment 2 procedure. This limitation should be addressed in future research by finding the appropriate balance in the trade-off between collecting lengthy survey data yet fatiguing subjects or collecting less data while preserving enthusiasm for the online experience. An additional limitation is the use of rank scores as primary measures

in the Study Strategy Survey. Without absolute measures such as number of hours spent engaged in a particular strategy, it is difficult to determine whether a strategy rank has changed due to decreased priority of a strategy previously preferable to it or due to an increased appreciation of that strategy itself. Furthermore, many educators might be interested in finding methods of increasing the amount of time that students spent engaged in a particular strategy, and care less about that strategy's perceived benefits. Finally, as previously noted, some students were confused by the wording on the study strategy survey (e.g., providing the same rank to multiple strategies which had to be manually corrected), so an improved version of the SSS is clearly desirable.

To conclude, the present investigation has demonstrated that students will alter their beliefs about the relative benefits of study strategies and reprioritize the use of those strategies if exposed to the proper educational intervention that includes an in-person lecture component. Although weaker in magnitude, significant changes to attitudes and behaviors related to study strategies are observed even in the absence of in-person lecture. These benefits are especially relevant to the direct comparative experience of the testing effects as well as goal and implementation intentions (specific to the experimental intervention) recommended by the Model of Action Phases (Bamberg, 2013; Heckhausen & Gollwitzer, 1987). Future interventions should now be developed to surpass the benchmarks of study strategy reprioritization elicited by Quiz-Me.

REFERENCES

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659-701.

Bamberg, S. (2013). Changing environmentally harmful behaviors: A stage model of self-regulated behavioral change. *Journal of Environmental Psychology, 34*, 151-159.

Belluck, P. (2011, January 20). *To really learn, quit studying and take a test.* New York Times. Retrieved from https://www.nytimes.com/2011/01/21/science/21memory.html.

de Winstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition, 32*, 945-955.

Dobson, J. L. & Linderholm, T. (2015). Self-testing promotes superior retention of anatomy and physiology information. *Advances in Health Science Education, 20,* 149-161.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*, 4-58.

Duncan, T. G., & McKeachie, W. J. (2005). The Making of the motivated strategies for learning questionnaire, *Educational Psychologist*, *40*, 117-128.

Einstein, G. O., Mullet, H. G., Harrison, T. L. (2012). The Testing effect: Illustrating a fundamental concept and changing study strategies. *Teaching of Psychology, 39,* 190-193.

Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology, 4,* 195-202.

Freemark, S. (2014, October 16). *Studying with quizzes helps make sure the material sticks.* KQED News. Retrieved from https://www.kqed.org/mindshift/37752/studying-with-quizzes-helps-make-sure-the-material-sticks.

Geller, J., Toftness, A. R., Armstrong, P. I., Carpenter, S. K., Manz, C. L., Coffman, C. R., & Lamm, M. H. (2017). Study strategies and beliefs about learning as a function of academic achievement and achievement goals. *Memory, 26*, 683-690.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology, Vol 38* (pp. 69-119). Elsevier Academic Press, San Diego, CA.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review, 19*(1), 126-134.

Heckhausen, H., & Gollwitzer, P. M. (1987). Thought contents and cognitive functioning in motivational versus volitional states of mind. *Motivation and Emotion, 11*, 101-120.

Karpicke, J. D., Butler, A. C., & Roediger, H. L., III. (2009). Metacognitive strategies in student learning: Do students practise retrieval when they study on their own? *Memory, 17*, 471-479.

Karpicke, J. D., & Zaromb, F. M. (2010). Retrieval model distinguishes the testing effect from the generation effect. *Journal of Memory and Language, 62,* 227-239.

Knäuper, B., Roseman, M., Johnson, P. J., & Krantz, L. H. (2009). Using mental imagery to enhance the effectiveness of implementation intentions. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues, 28*, 181-186.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*, 219-224.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language, 65,* 85-97.

McAndrew, M., Morrow, C. S., Atiyeh, L., & Pierre, G. C. (2016). Dental student study strategies: Are self-testing and scheduling related to academic performance? *Journal of Dental Education,* 80, 542–552.

McCrea, S. M., Penningroth, S. L., & Radakovich, M. P. (2015). Implementation intentions forge a strong cue–response link and boost prospective memory performance. *Journal of Cognitive Psychology, 27*, 12-26.

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

Meslot, C., Gauchet, A., Allenet, B., François, O., & Hagger, M. S. (2016). Theory-based interventions combining mental simulation and planning techniques to improve physical activity: Null results from two randomized controlled trials. *Frontiers in Psychology, 7*, 1-17.

Pan, S. C,. & Rickard, T. C. (in press). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*.

Pakpour, A. H., Ghloami, M., Gellert, P., Tekaninejad, M. S., Dombrowski, S. U., & Webb, T. L. (2016). The effects of two planning interventions on the oral health behavior of Iranian adolescents: A cluster randomized controlled trial. *Annals of Behavioral Medicine, 50,* 409-418.

Pintrich, P. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82,* 33-40.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement, 53,* 801-813.

Roediger, H. L., III, Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied, 17*, 382-395.

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

Roediger, H. L., III, & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210.

Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 233-239.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140*, 1432-1463.

Shapiro, A. M., & Gordon, L. T. (2012). A controlled study of clicker-assisted memory enhancement in college classrooms. *Applied Cognitive Psychology, 26*, 635-643.

Sheeran, P. (2002). Intention-behavior relations: A conceptual and empirical review. *European Review of Social Psychology, 12,* 1-36.

Sheeran, P., & Webb, T. L. (2016). The intention–behavior gap. *Social and Personality Psychology Compass, 10*, 503-518.

Tack, L. A., Healy, A. F., & Jones, M. (2014, June). *Isolating the effects of different feedback contents on learning*. Poster presented at the Festschrift in Honor of Alice F. Healy, Boulder, CO.

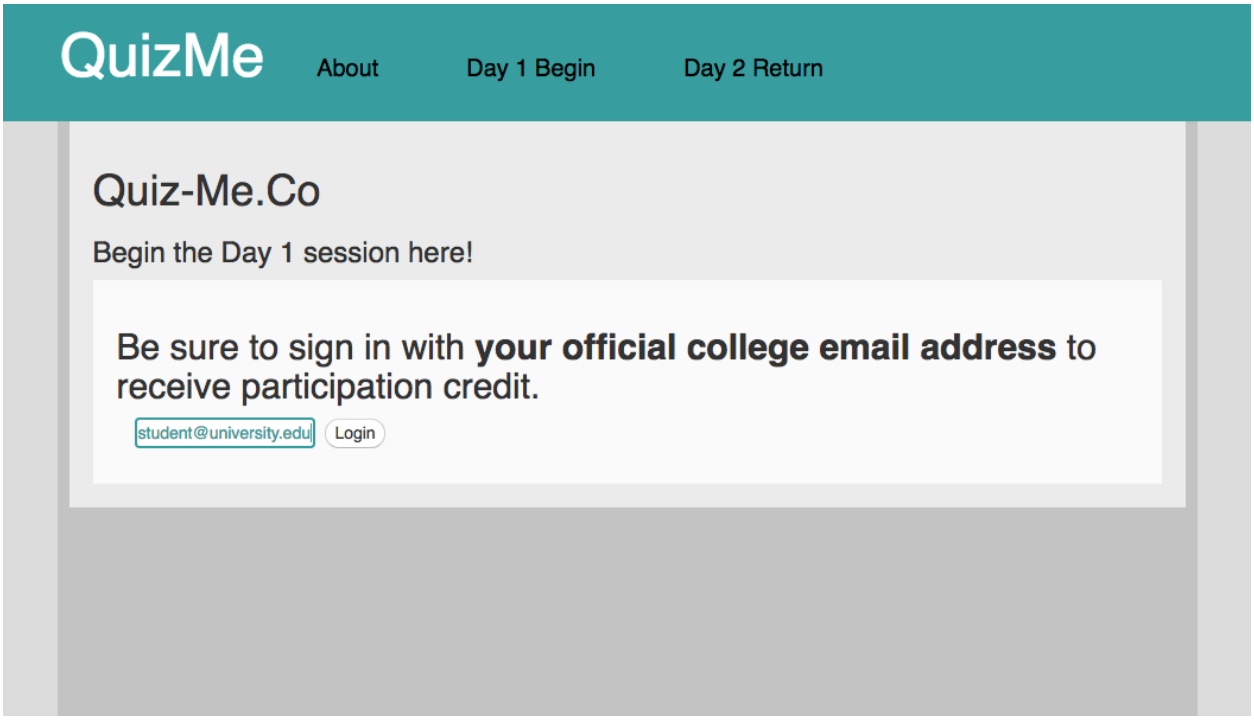
Yan, V. X., Thai, K., & Bjork, R. A. (2014). Habits and beliefs that guide self-regulated learning: Do they vary with mindset? *Journal of Applied Research in Memory and Cognition, 3*(3), 140-152.

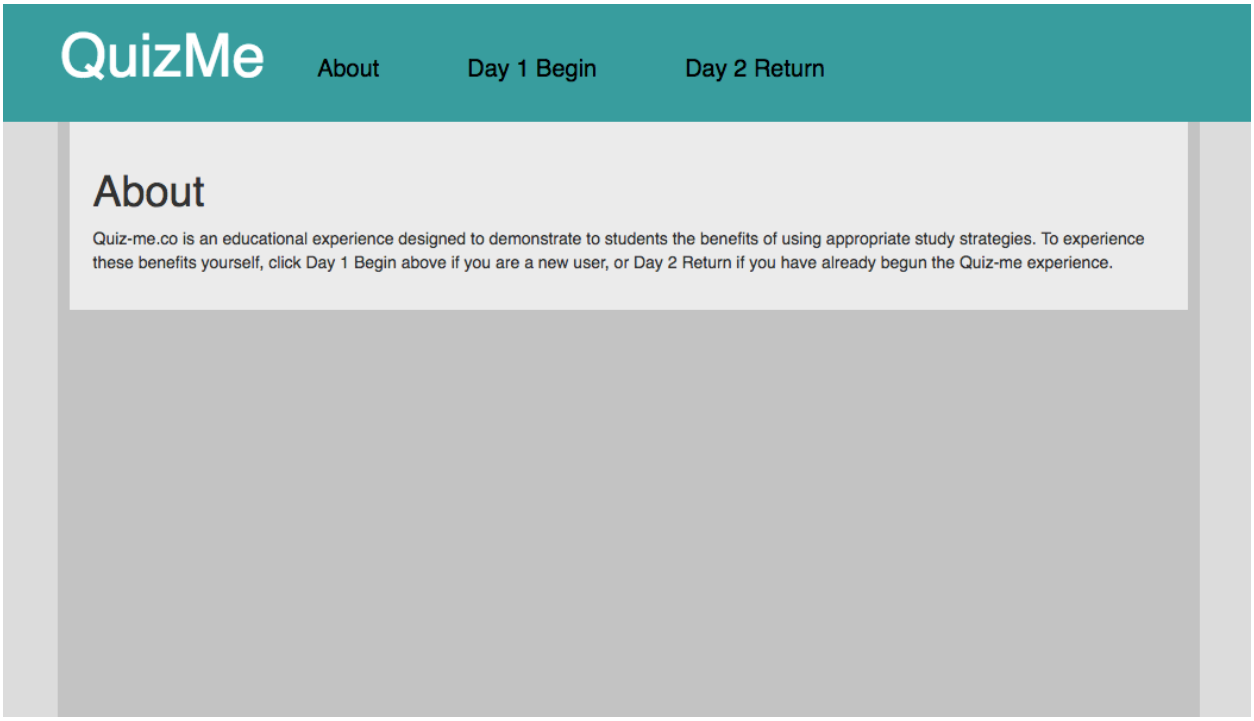
Young, A. P., Healy, A. F., Jones, M., & Curran, T. (2016, April). *Comparative Experience is Not Enough: Challenges of Eliciting Awareness of the Testing Effect*. Invited paper presented in the symposium Studies from the Center for Research on Training, University of Colorado Boulder at the 86th Annual Rocky Mountain Psychological Association Convention. Denver, CO.

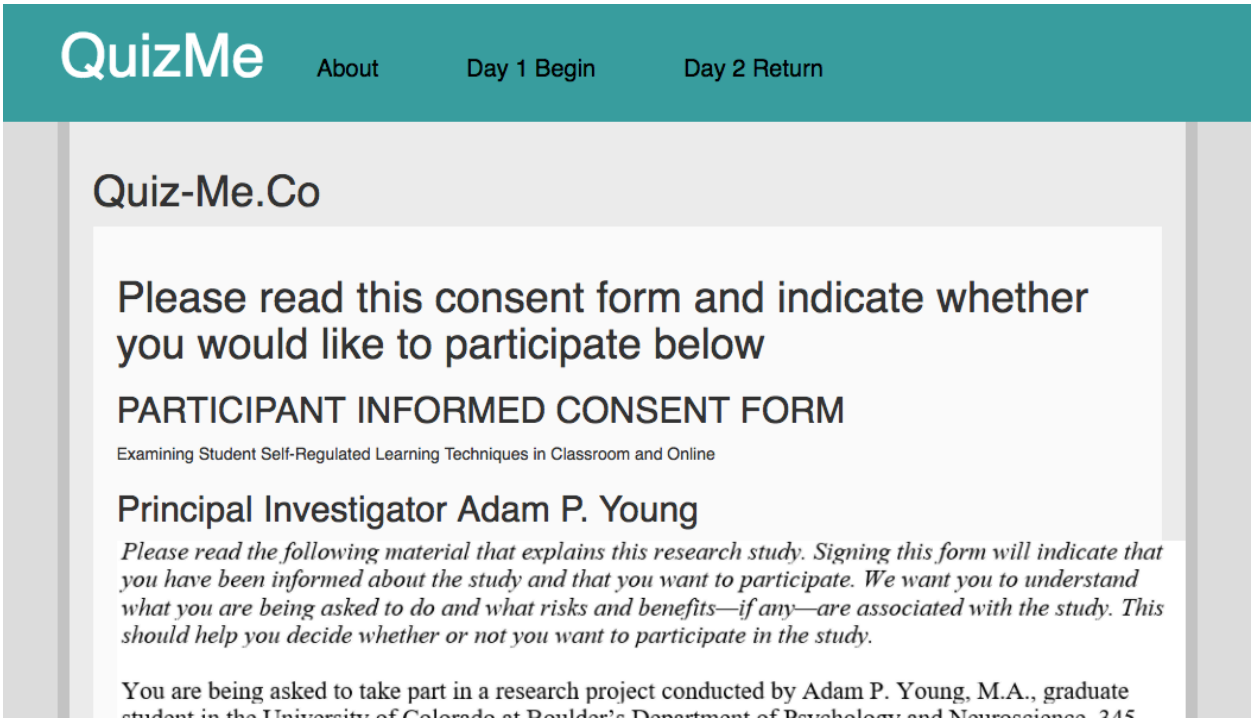
Appendix A:

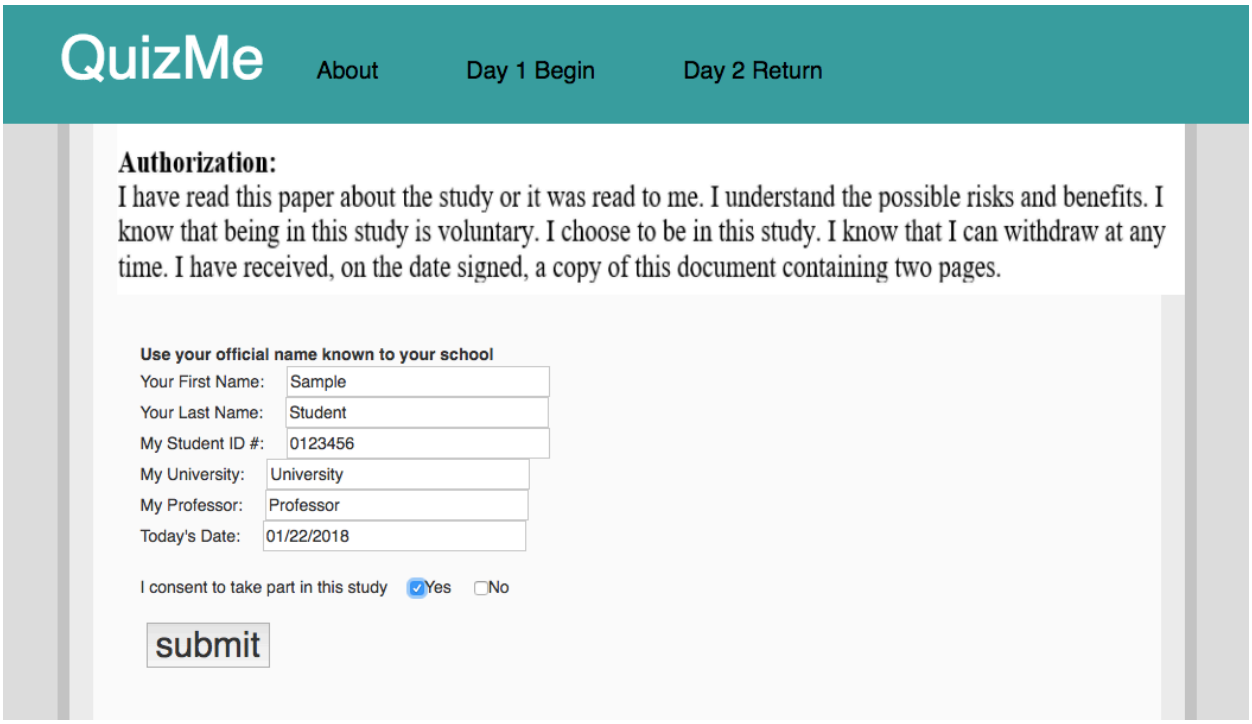
Images of the Final Quiz-Me Online Intervention (Spring 2018)

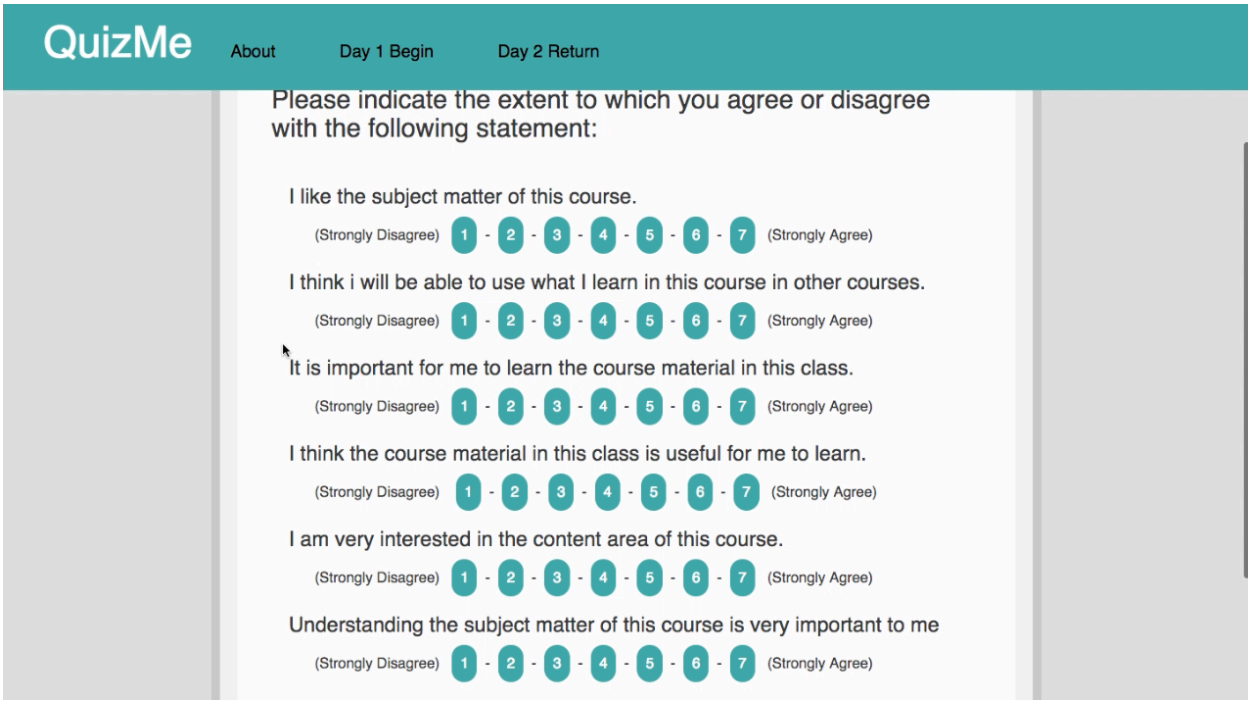
See text for divergences from Fall 2017 version

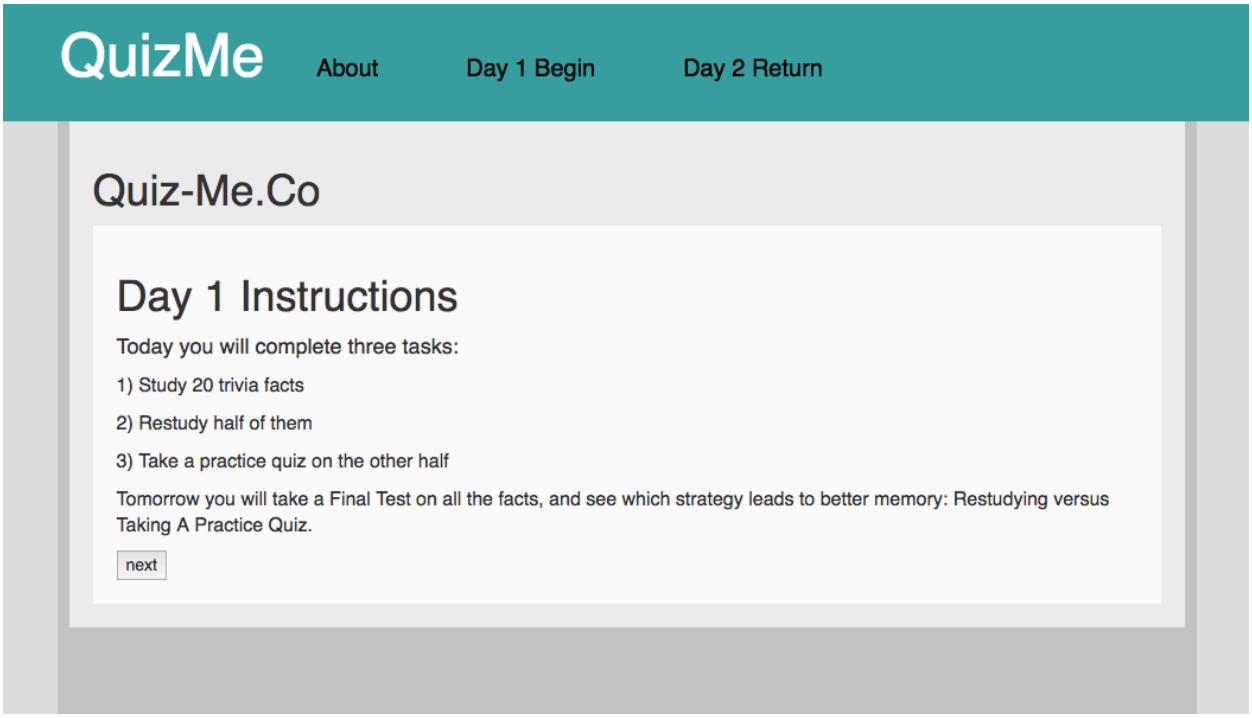


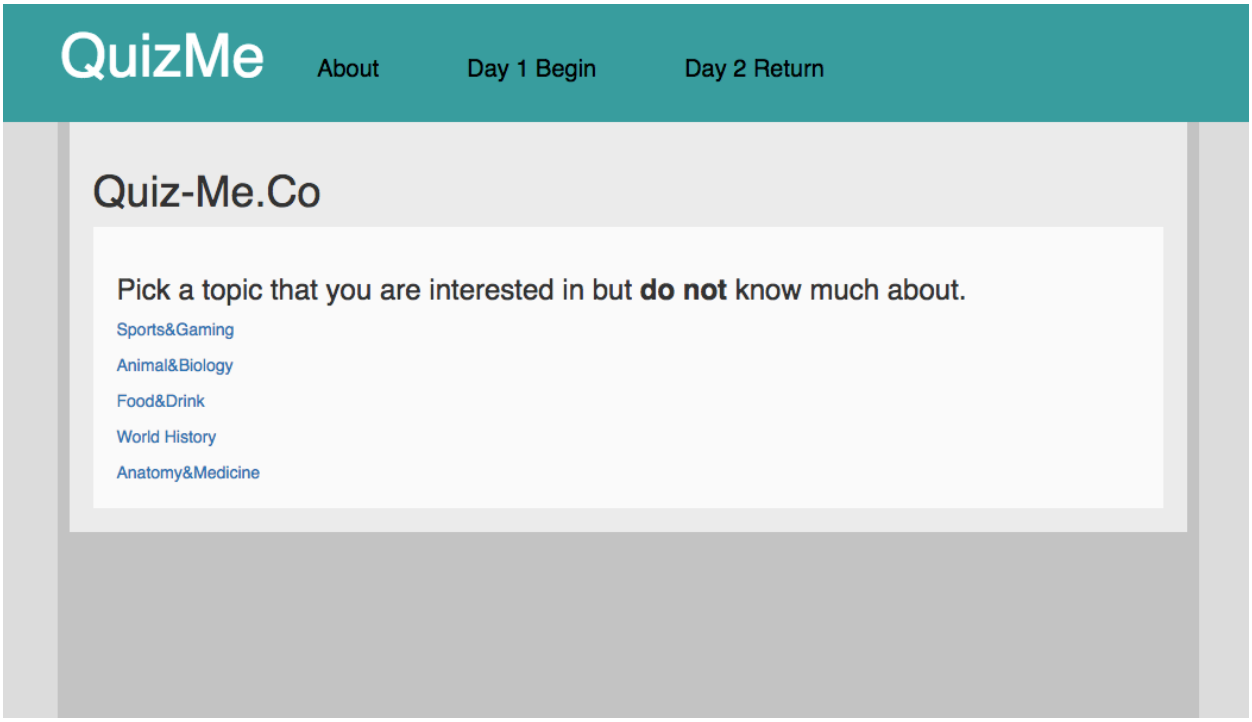


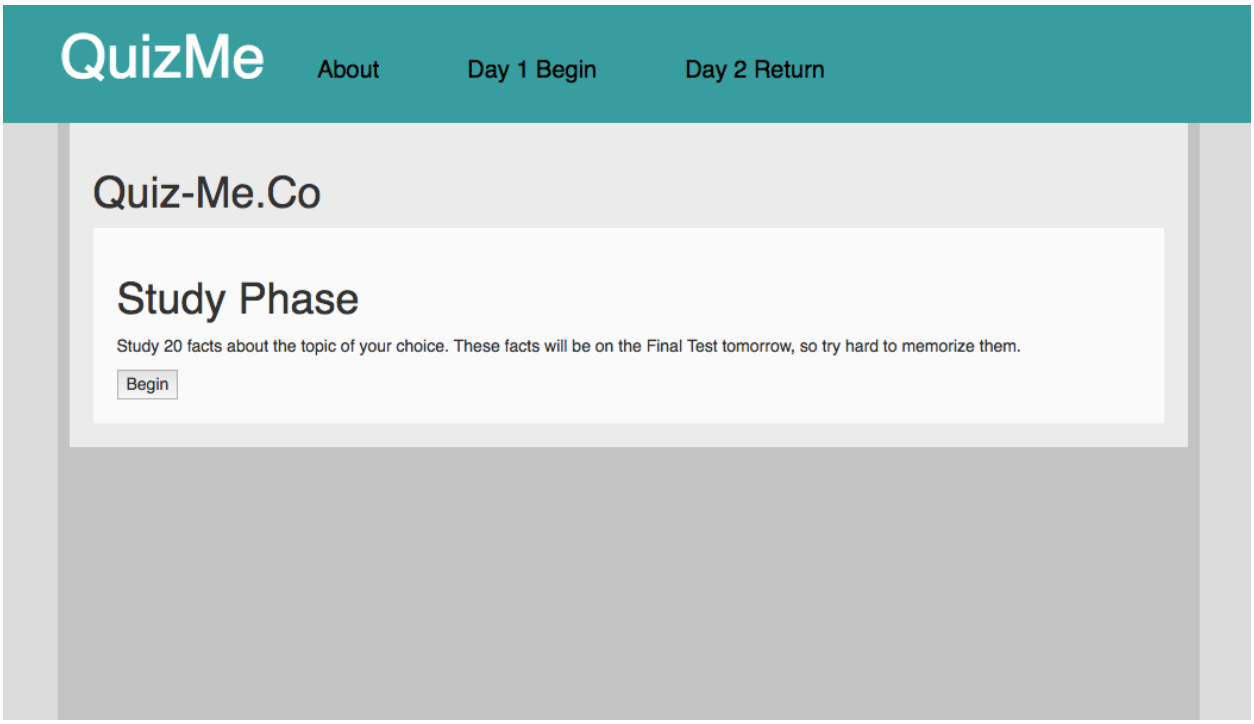

QuizMe
About
Day 1 Begin
Day 2 Return
Quiz-Me.Co
Please read this consent form and indicate whether you would like to participate below
PARTICIPANT INFORMED CONSENT FORM
Examining Student Self-Regulated Learning Techniques in Classroom and Online
Principal Investigator Adam P. Young
Please read the following material that explains this research study. Signing this form will indicate that you have been informed about the study and that you want to participate. We want you to understand what you are being asked to do and what risks and benefits—if any—are associated with the study. This should help you decide whether or not you want to participate in the study.
You are being asked to take part in a research project conducted by Adam P. Young, M.A., graduate

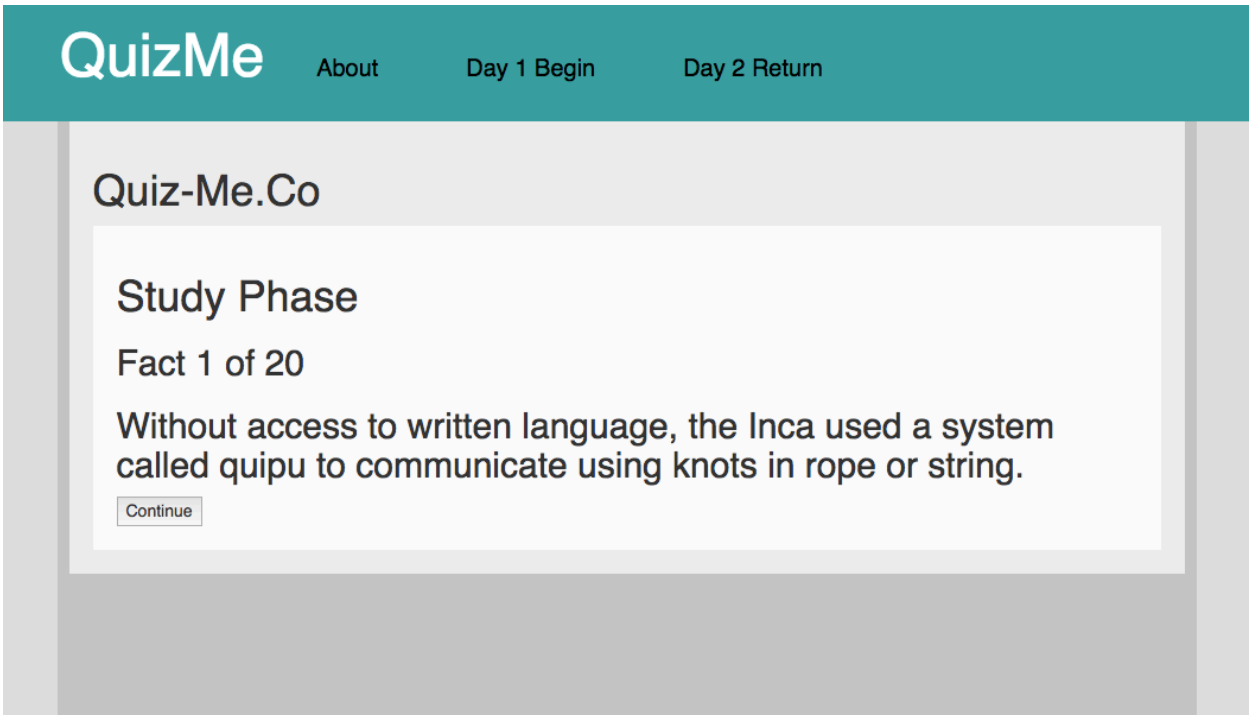

QuizMe
About
Day 1 Begin
Day 2 Return
Authorization:
I have read this paper about the study or it was read to me. I understand the possible risks and benefits. I know that being in this study is voluntary. I choose to be in this study. I know that I can withdraw at any time. I have received, on the date signed, a copy of this document containing two pages.
Use your official name known to your school
Your First Name: Sample
Your Last Name: Student
My Student ID #: 0123456
My University: University
My Professor: Professor
Today's Date: 01/22/2018
I consent to take part in this study Yes No
submit

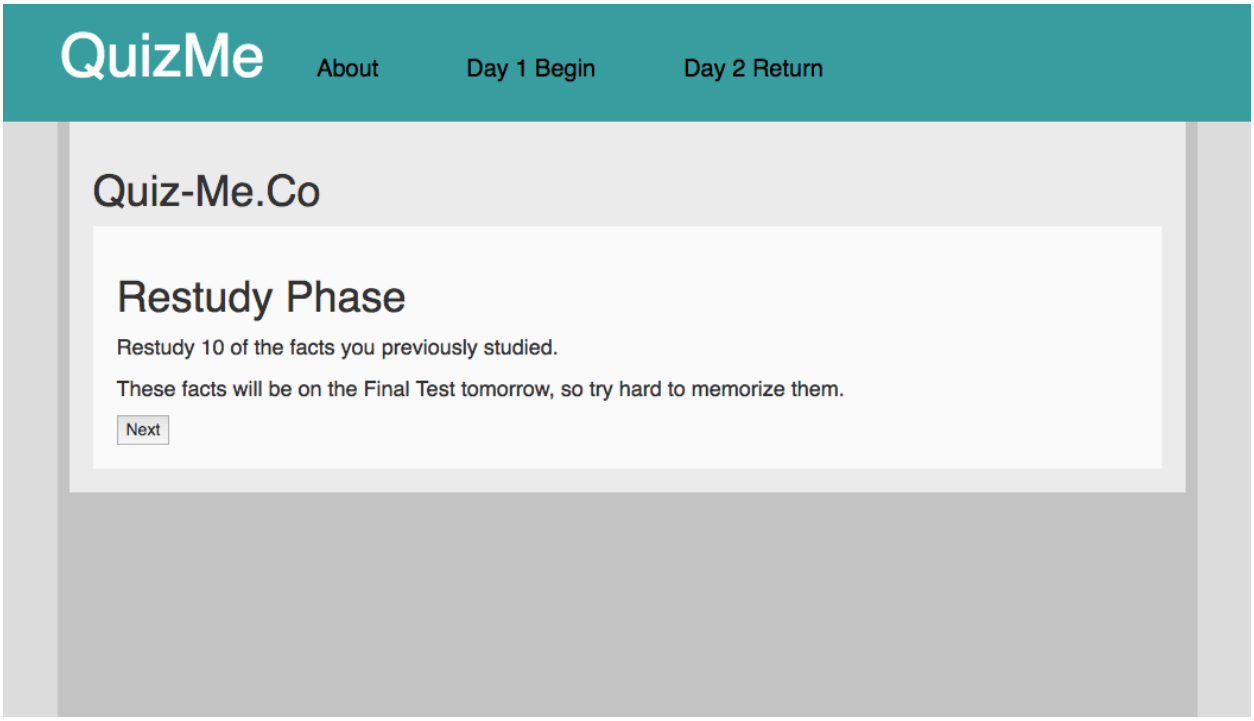

QuizMe
About
Day 1 Begin
Day 2 Return
Please indicate the extent to which you agree or disagree with the following statement:
I like the subject matter of this course.
(Strongly Disagree) 1 - 2 - 3 - 4 - 5 - 6 - 7 (Strongly Agree)
I think i will be able to use what I learn in this course in other courses.
(Strongly Disagree) 1 - 2 - 3 - 4 - 5 - 6 - 7 (Strongly Agree)
It is important for me to learn the course material in this class.
(Strongly Disagree) 1 - 2 - 3 - 4 - 5 - 6 - 7 (Strongly Agree)
I think the course material in this class is useful for me to learn.
(Strongly Disagree) 1 - 2 - 3 - 4 - 5 - 6 - 7 (Strongly Agree)
I am very interested in the content area of this course.
(Strongly Disagree) 1 - 2 - 3 - 4 - 5 - 6 - 7 (Strongly Agree)
Understanding the subject matter of this course is very important to me
(Strongly Disagree) 1 - 2 - 3 - 4 - 5 - 6 - 7 (Strongly Agree)

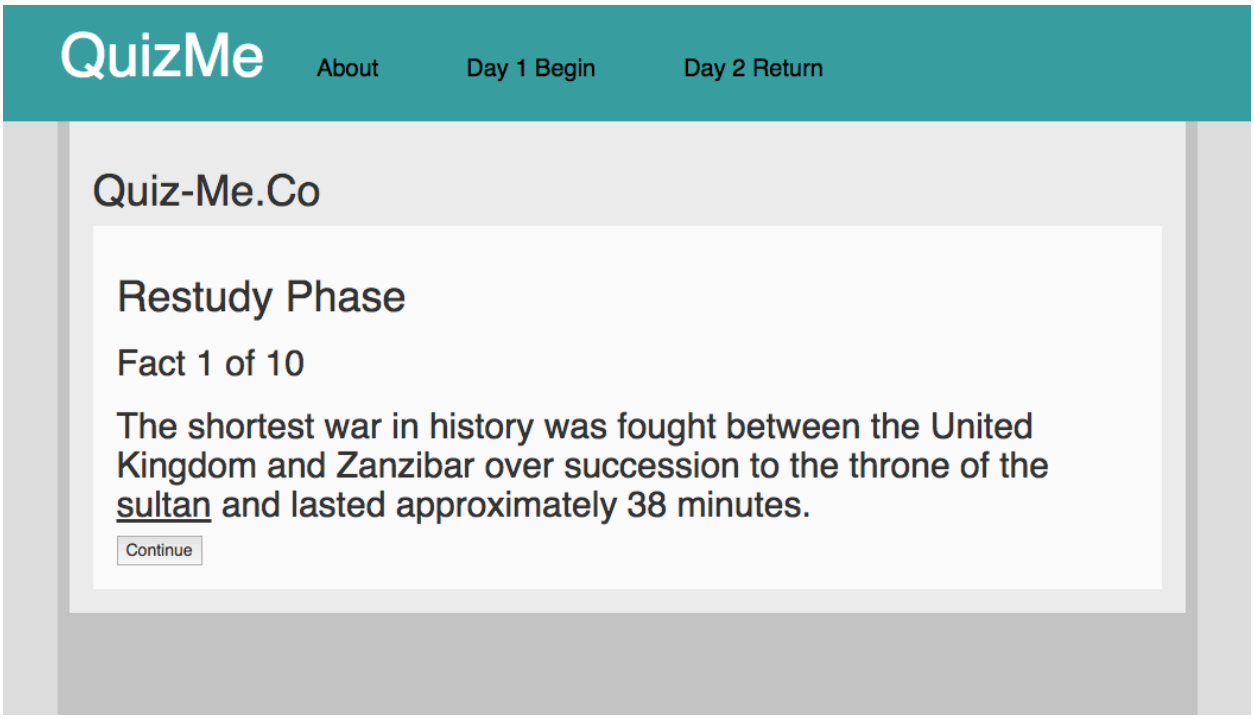

QuizMe
About
Day 1 Begin
Day 2 Return
Quiz-Me.Co
Day 1 Instructions
Today you will complete three tasks:
1) Study 20 trivia facts
2) Restudy half of them
3) Take a practice quiz on the other half
Tomorrow you will take a Final Test on all the facts, and see which strategy leads to better memory: Restudying versus Taking A Practice Quiz.
next

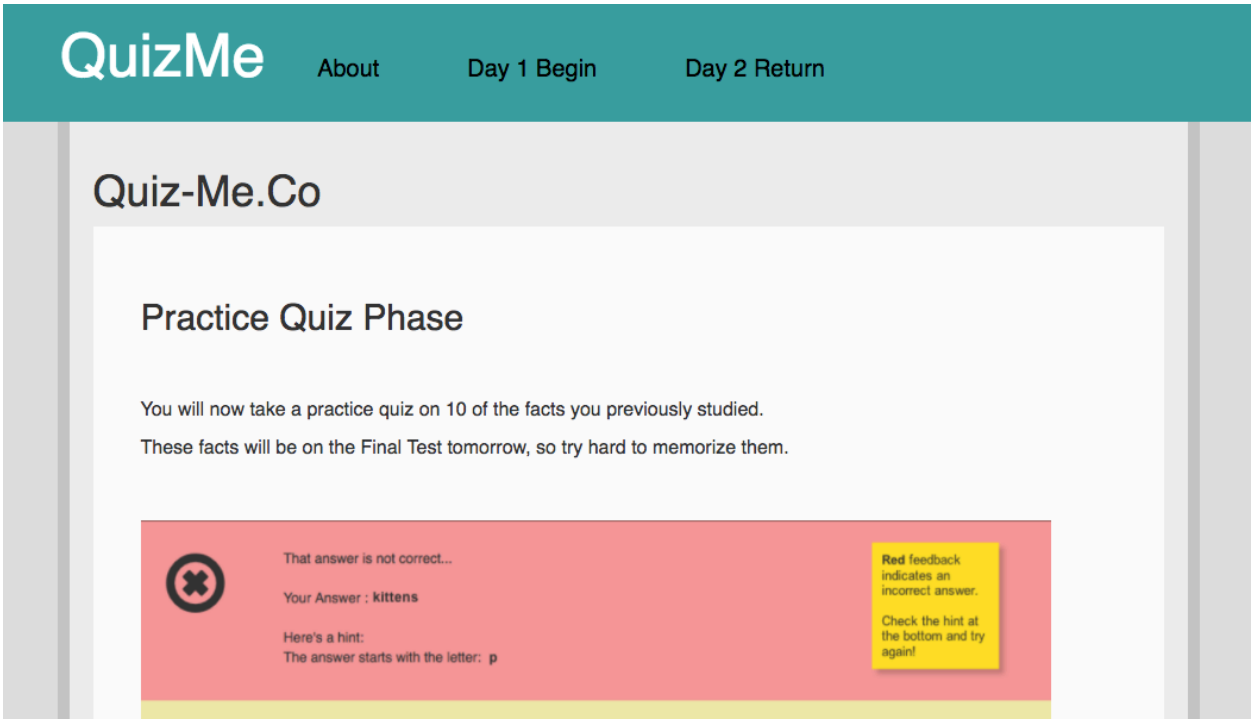

QuizMe
About
Day 1 Begin
Day 2 Return
Quiz-Me.Co
Pick a topic that you are interested in but **do not** know much about.
Sports&Gaming
Animal&Biology
Food&Drink
World History
Anatomy&Medicine


QuizMe
About
Day 1 Begin
Day 2 Return
Quiz-Me.Co
Study Phase
Study 20 facts about the topic of your choice. These facts will be on the Final Test tomorrow, so try hard to memorize them.
Begin

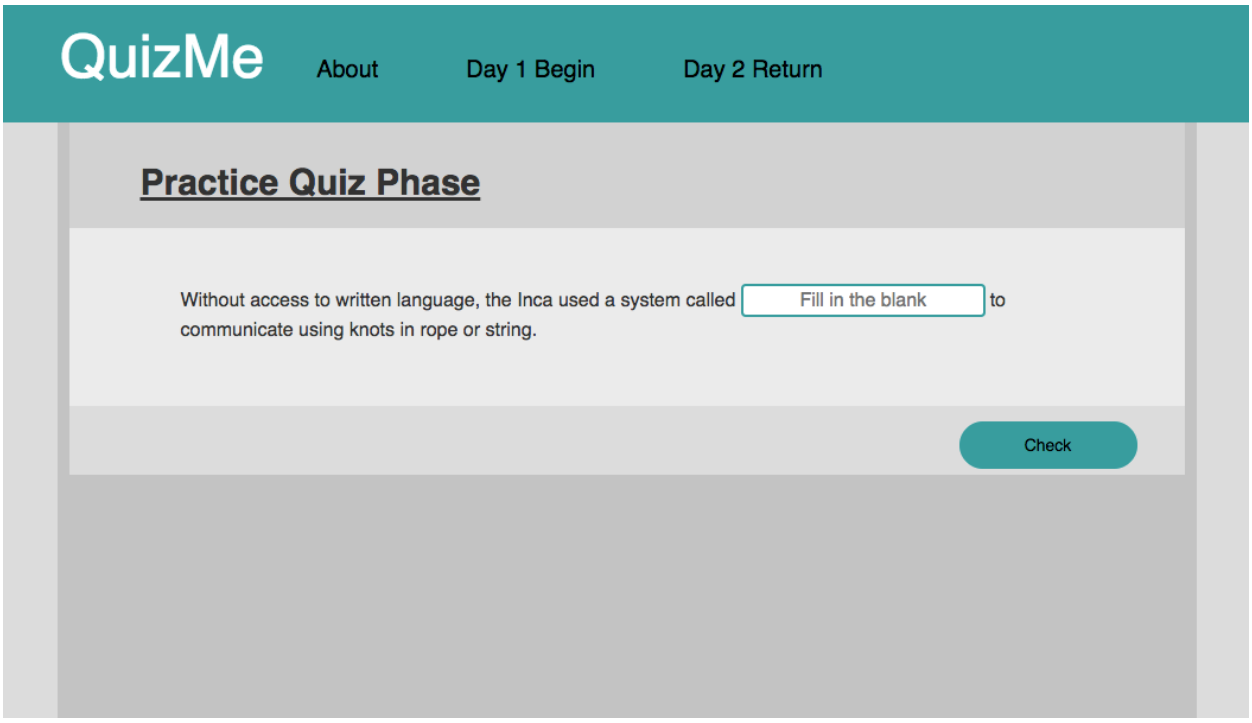

QuizMe
About
Day 1 Begin
Day 2 Return
Quiz-Me.Co
Study Phase
Fact 1 of 20
Without access to written language, the Inca used a system called quipu to communicate using knots in rope or string.
Continue

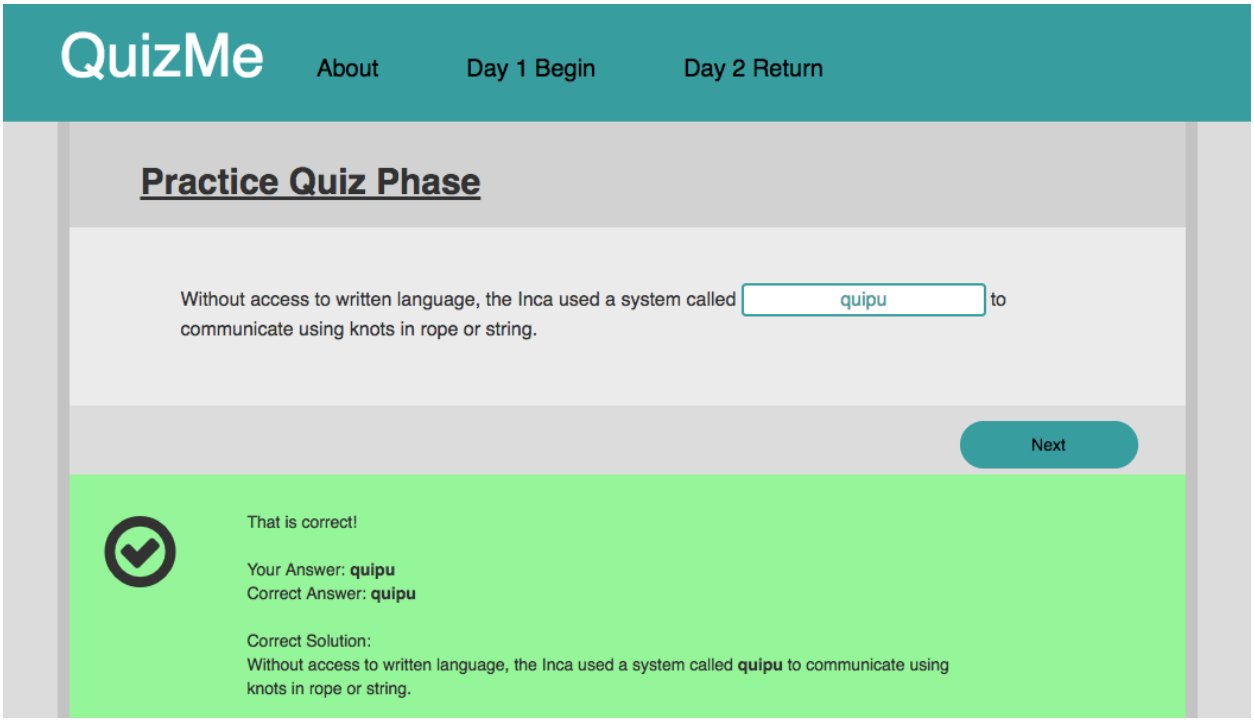

QuizMe
About
Day 1 Begin
Day 2 Return
Quiz-Me.Co
Restudy Phase
Restudy 10 of the facts you previously studied.
These facts will be on the Final Test tomorrow, so try hard to memorize them.
Next

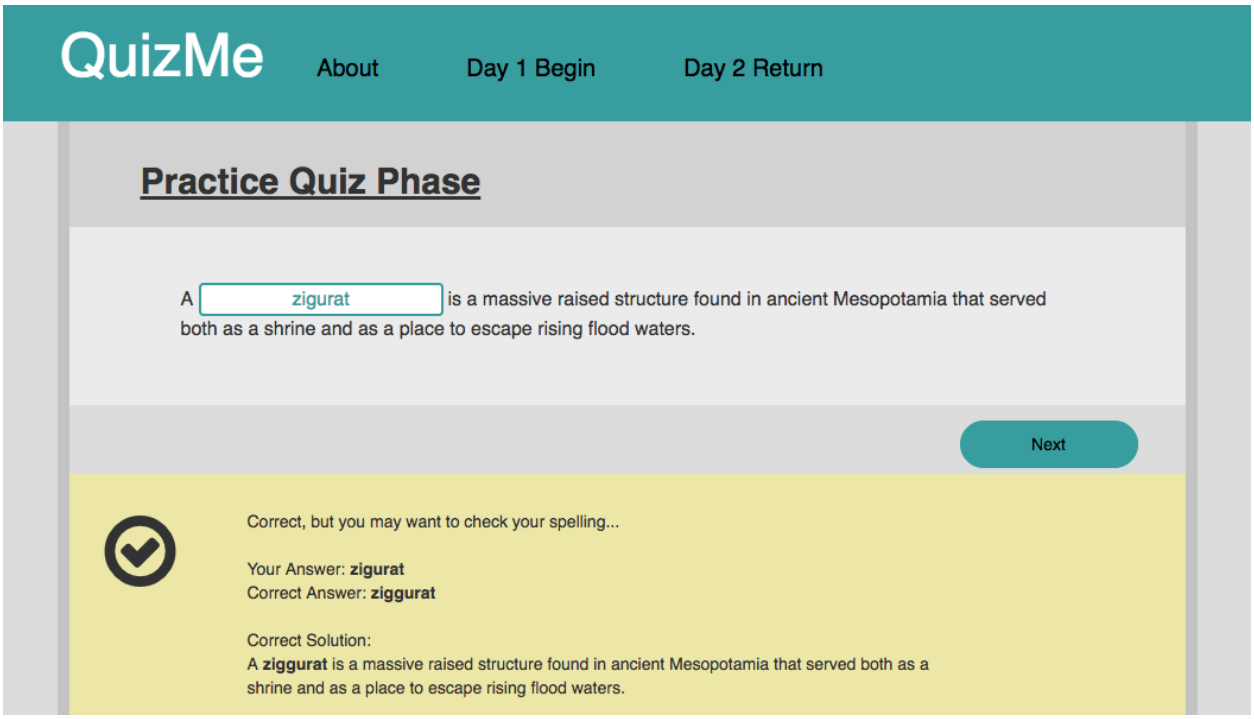

QuizMe
About
Day 1 Begin
Day 2 Return
Quiz-Me.Co
Restudy Phase
Fact 1 of 10
The shortest war in history was fought between the United Kingdom and Zanzibar over succession to the throne of the sultan and lasted approximately 38 minutes.
Continue

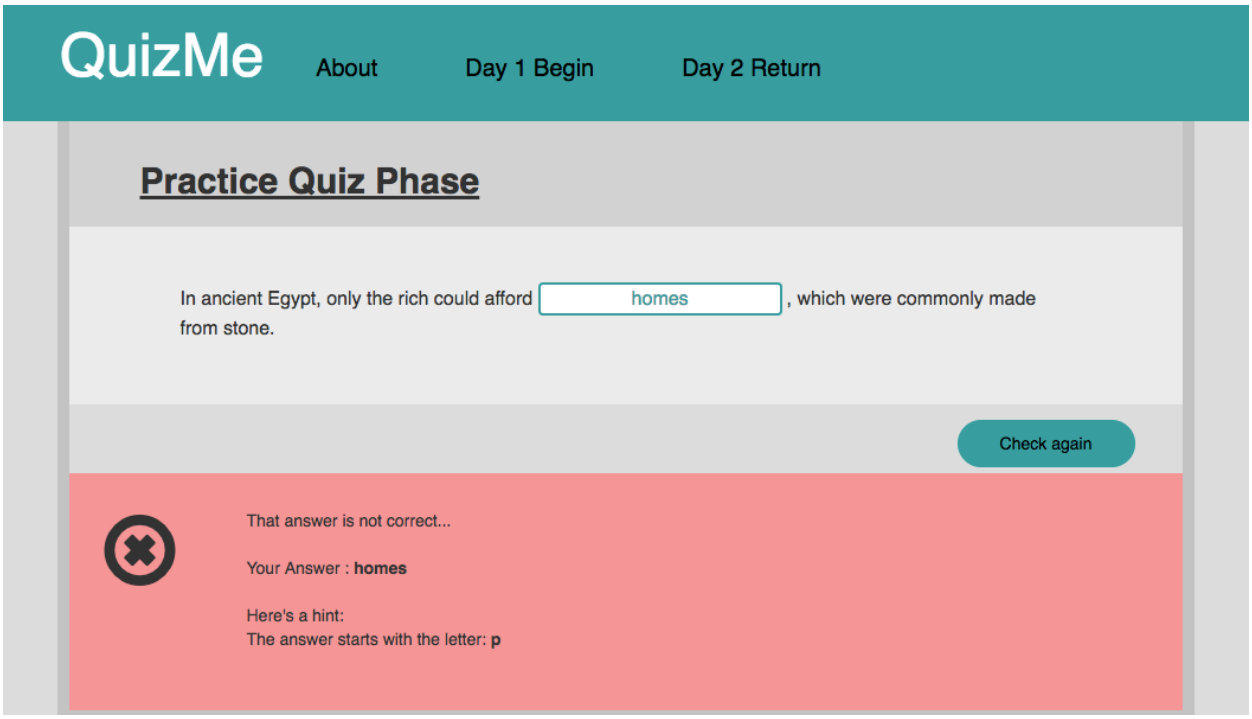

QuizMe
About
Day 1 Begin
Day 2 Return
Quiz-Me.Co
Practice Quiz Phase
You will now take a practice quiz on 10 of the facts you previously studied.
These facts will be on the Final Test tomorrow, so try hard to memorize them.
That answer is not correct...
Your Answer : kittens
Here's a hint:
The answer starts with the letter: p
Red feedback indicates an incorrect answer.
Check the hint at the bottom and try again!

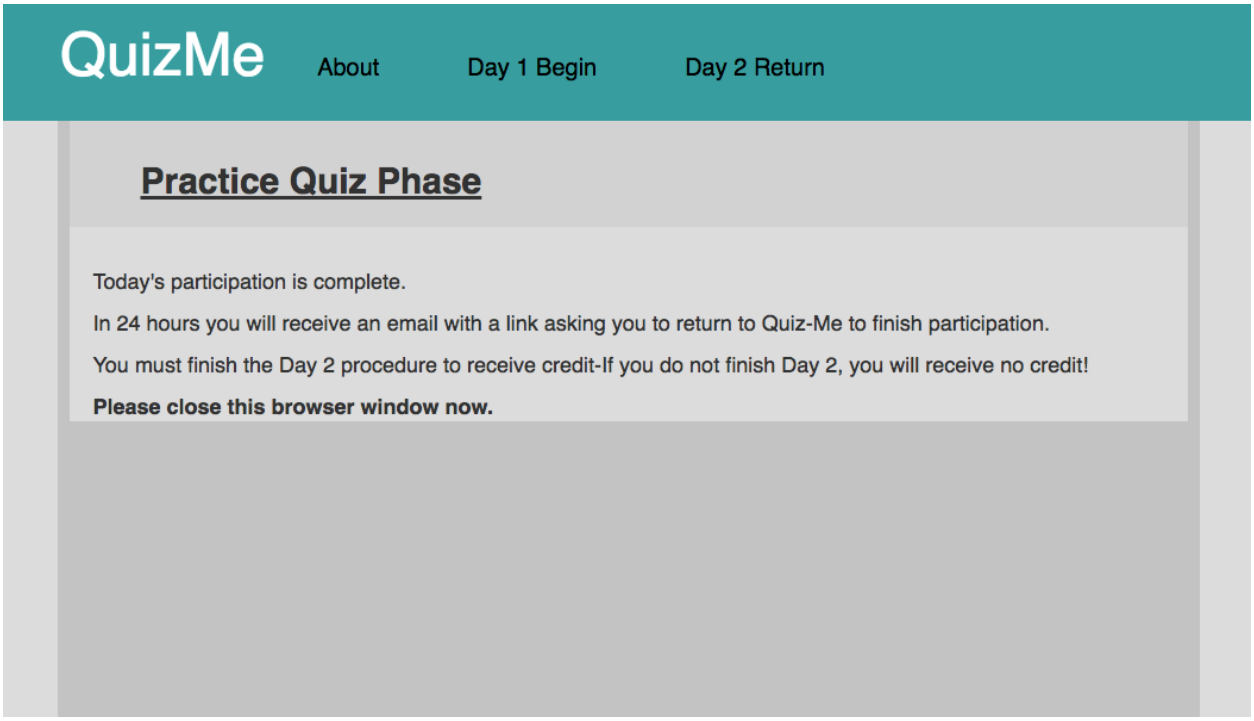

QuizMe
About
Day 1 Begin
Day 2 Return
That answer is not correct...
Your Answer : kittens
Here's a hint:
The answer starts with the letter: p
Red feedback indicates an incorrect answer.
Check the hint at the bottom and try again!
Correct, but you may want to check your spelling...
Your Answer: puppys
Correct Answer: puppies
Correct Solution: There are more than 5 million puppies born in the United States every day.
Yellow feedback indicates a correct answer that was spelled incorrectly.
That is correct!
Your Answer: puppies
Correct Answer: puppies
Correct Solution: There are more than 5 million puppies born in the United States every day.
Green feedback indicates a correct answer that was spelled correctly.
Nice job!

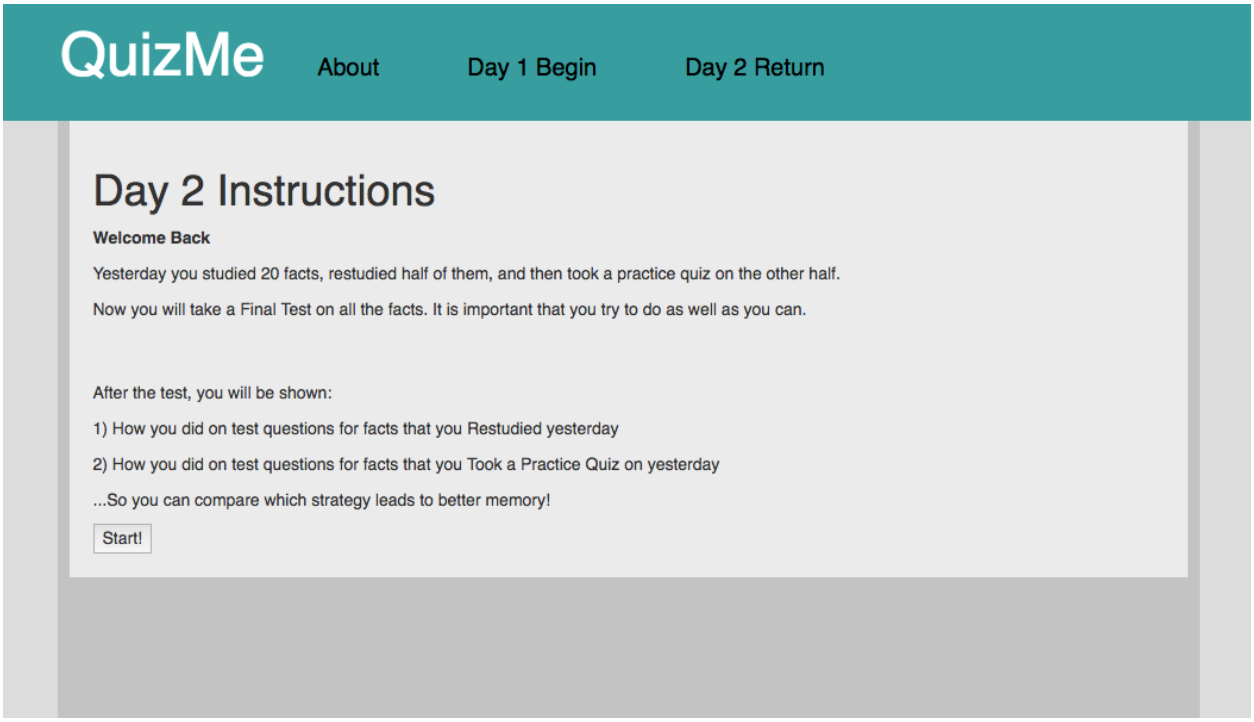

QuizMe
About
Day 1 Begin
Day 2 Return
Practice Quiz Phase
Without access to written language, the Inca used a system called Fill in the blank to communicate using knots in rope or string.
Check

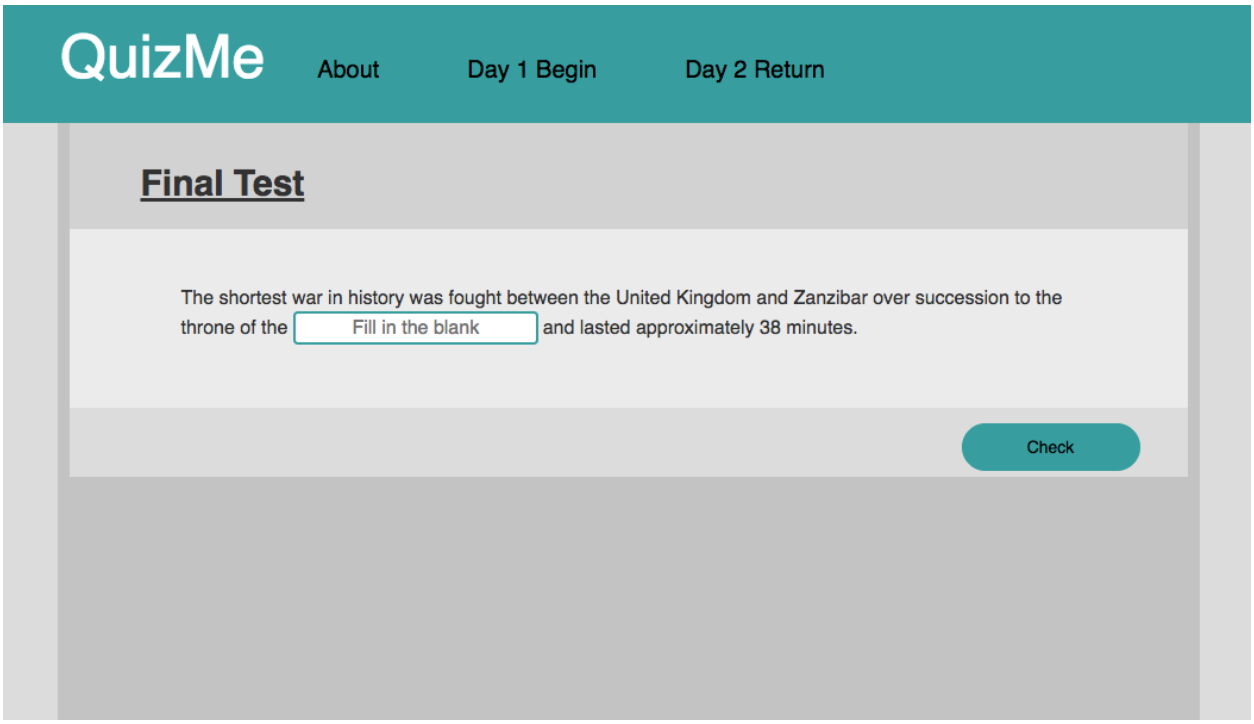

QuizMe
About
Day 1 Begin
Day 2 Return
Practice Quiz Phase
Without access to written language, the Inca used a system called quipu to communicate using knots in rope or string.
Next
That is correct!
Your Answer: quipu
Correct Answer: quipu
Correct Solution:
Without access to written language, the Inca used a system called quipu to communicate using knots in rope or string.

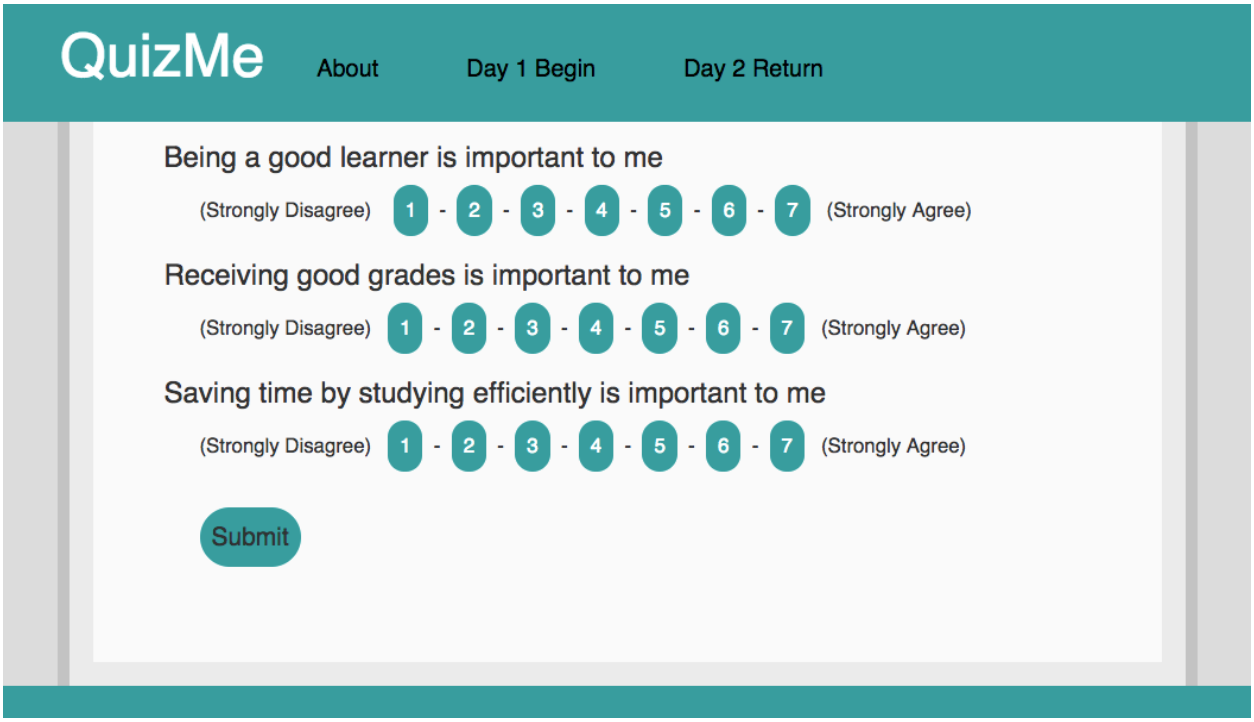

QuizMe
About
Day 1 Begin
Day 2 Return
Practice Quiz Phase
A zigurat is a massive raised structure found in ancient Mesopotamia that served both as a shrine and as a place to escape rising flood waters.
Next
Correct, but you may want to check your spelling...
Your Answer: zigurat
Correct Answer: ziggurat
Correct Solution:
A ziggurat is a massive raised structure found in ancient Mesopotamia that served both as a shrine and as a place to escape rising flood waters.

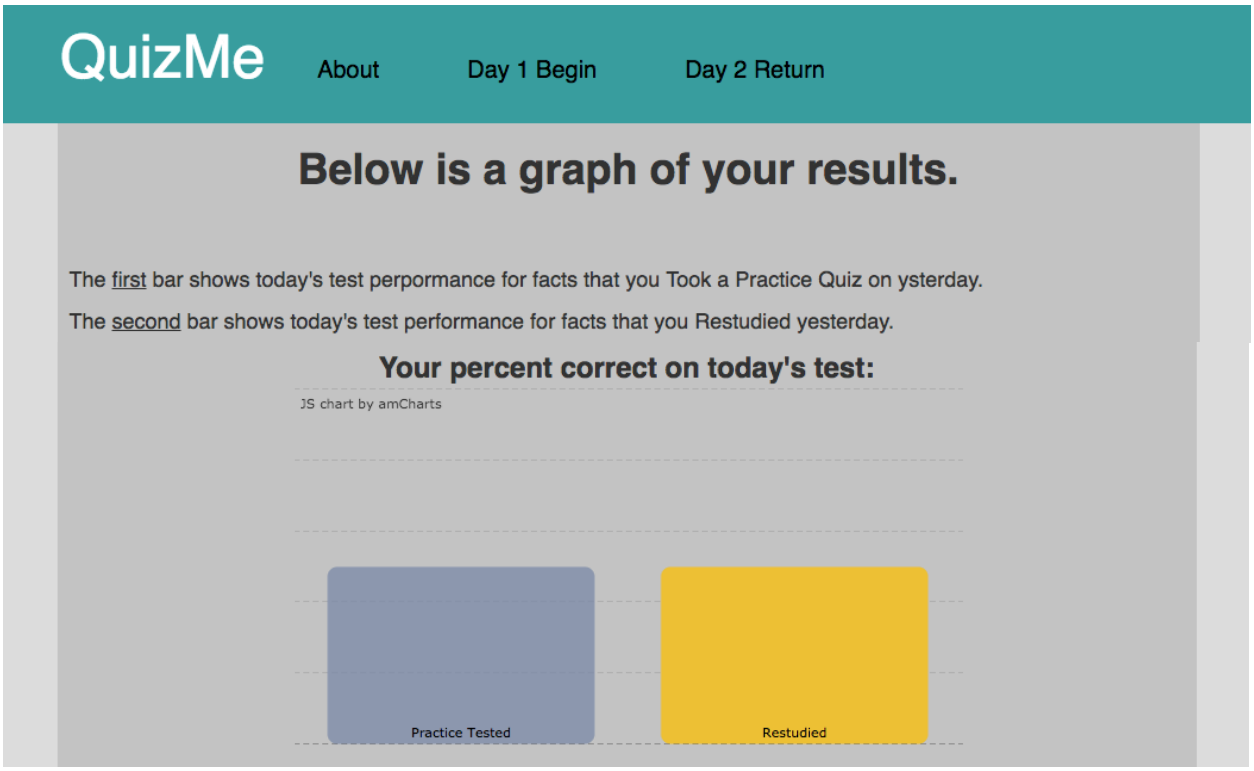

QuizMe
About
Day 1 Begin
Day 2 Return
Practice Quiz Phase
In ancient Egypt, only the rich could afford homes , which were commonly made from stone.
Check again
That answer is not correct...
Your Answer : homes
Here's a hint:
The answer starts with the letter: p


QuizMe
About
Day 1 Begin
Day 2 Return
Practice Quiz Phase
Today's participation is complete.
In 24 hours you will receive an email with a link asking you to return to Quiz-Me to finish participation.
You must finish the Day 2 procedure to receive credit-If you do not finish Day 2, you will receive no credit!
Please close this browser window now.


QuizMe
About
Day 1 Begin
Day 2 Return
Day 2 Instructions
Welcome Back
Yesterday you studied 20 facts, restudied half of them, and then took a practice quiz on the other half.
Now you will take a Final Test on all the facts. It is important that you try to do as well as you can.
After the test, you will be shown:
1) How you did on test questions for facts that you Restudied yesterday
2) How you did on test questions for facts that you Took a Practice Quiz on yesterday
...So you can compare which strategy leads to better memory!
Start!


QuizMe
About
Day 1 Begin
Day 2 Return
Final Test
The shortest war in history was fought between the United Kingdom and Zanzibar over succession to the throne of the
Fill in the blank
and lasted approximately 38 minutes.
Check


QuizMe
About
Day 1 Begin
Day 2 Return
Being a good learner is important to me
(Strongly Disagree) 1 - 2 - 3 - 4 - 5 - 6 - 7 (Strongly Agree)
Receiving good grades is important to me
(Strongly Disagree) 1 - 2 - 3 - 4 - 5 - 6 - 7 (Strongly Agree)
Saving time by studying efficiently is important to me
(Strongly Disagree) 1 - 2 - 3 - 4 - 5 - 6 - 7 (Strongly Agree)
Submit

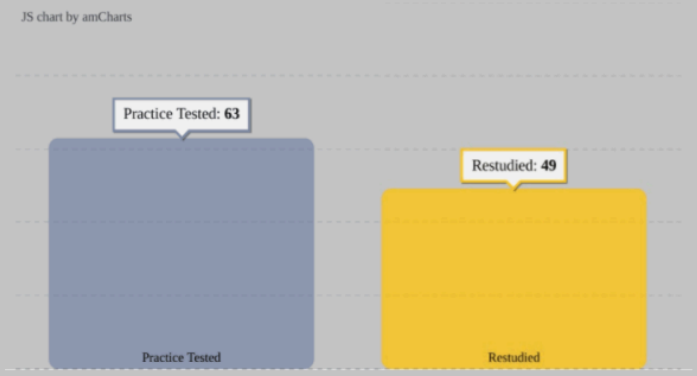

QuizMe
About
Day 1 Begin
Day 2 Return
Below is a graph of your results.
The first bar shows today's test perpormance for facts that you Took a Practice Quiz on ysterday.
The second bar shows today's test performance for facts that you Restudied yesterday.
Your percent correct on today's test:
JS chart by amCharts
Practice Tested
Restudied

Below is an average students' result.

The first bar shows how the average student did on facts that they Took a Practice Quiz on yesterday.

The second bar shows how the average student did on facts that they Restudied yesterday.

Average students' percent correct on this test:



As you can see, the average student has **much better memory after taking practice tests** compared to only studying. If you value being an effective learner or receiving good class grades, this strategy is for you.

Three techniques to use practice tests for better memory:

1) **Use flashcards or Quizlet.com to test yourself**. Don't turn them over until you've tried to recall the answer from memory

2) **Quiz your friends and have them quiz you.** Try to stump each other, then review the correct answers

3) **See if you can explain diffcult concepts out loud or on paper.** Compare your explanation to the one from the book or your notes.

Which strategy above can you picture yourself using?

Why that one in particular?

Your comments here

Please state your new strategy in If-Then format.

For example, **If** you are in (the place and time you just just identified), **Then** you will engage in (the strategy you chose to use).

Your comments here

That concludes the Tutorial experience!

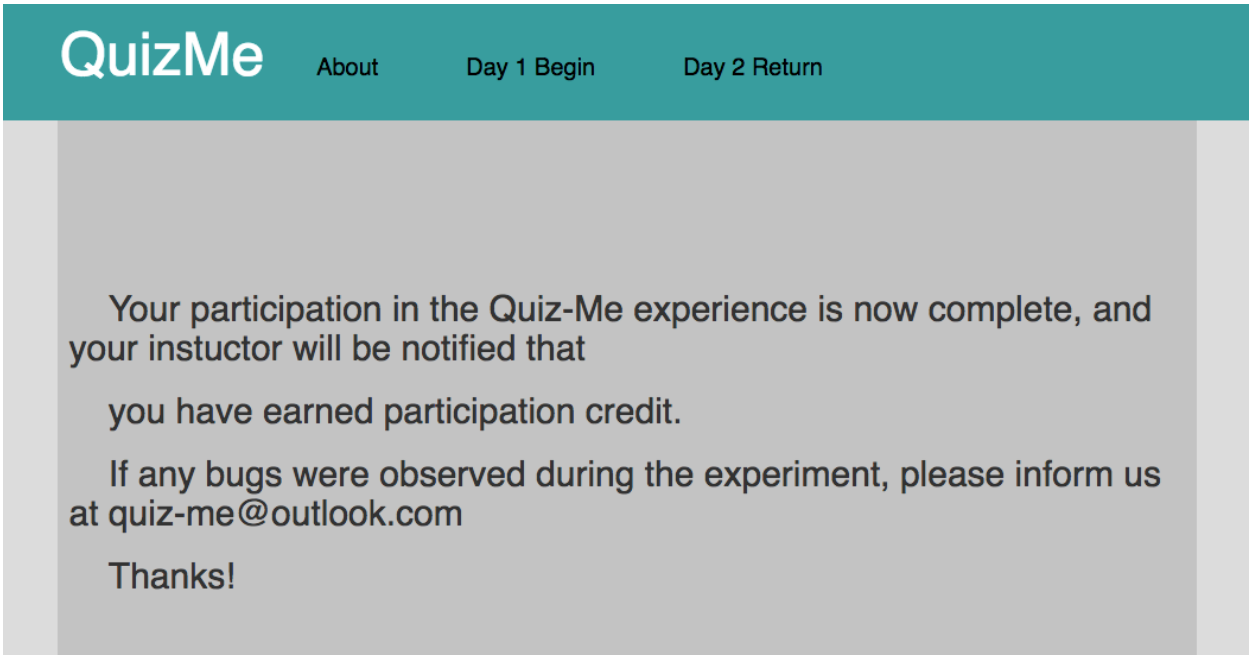
Unlike many other students, you now know the benefits of practice testing, and how to use it.

Thanks for participating, we hope it helps!

Do you believe you will use this information to help you study in the future?

Yes-I'll use this info

No-I won't use this info


QuizMe
About
Day 1 Begin
Day 2 Return
Your participation in the Quiz-Me experience is now complete, and your instuctor will be notified that
you have earned participation credit.
If any bugs were observed during the experiment, please inform us at quiz-me@outlook.com
Thanks!

Appendix B:

Motivated Strategies for Learning Questionnaire

And Cronbach's Alpha Reliability Values

Item	Scale	Alpha
Control of learning	It is my own fault if I don't learn the material in a course.	.663
	If I try hard enough, then I will understand the course material.	
	If I study in appropriate ways, then I will be able to learn the material I'm studying.	
	If I don't understand the course material, it is because I didn't try hard enough.	
Critical thinking	I try to play around with ideas of my own related to what I am learning in my courses.	.659
	I often find myself questioning things I hear or read in my courses to decide if I find them convincing.	
	Whenever I read or hear an assertion or conclusion in a class, I think about possible alternatives.	
	When a theory, interpretation, or conclusion is presented in class or in the readings, I try to decide if there is good supporting evidence.	
	I treat the course material as a starting point and try to develop my own ideas about it.	

Effort regulation	I work hard to do well in my classes even if I don't like what we are doing.	.820
	I often feel so lazy or bored when I study for a class that I quit before I finish what I planned to do. (REV)	
	When course work is difficult, I either give up or only study the easy parts. (REV)	
	Even when course materials are dull and uninteresting, I manage to keep working until I finish.	
Elaboration	When reading for a class, I try to relate the material to what I already know.	.617
	I try to relate ideas in one subject to those in other courses whenever possible.	
	I try to understand the material in a class by making connections between the readings and the concepts from the lectures.	
	When I study for a course, I write brief summaries of the main ideas from the readings and my class notes.	
	I try to apply ideas from course readings in other class activities such as lecture and discussion.	
	When I study for a class, I pull together information from different sources, such as lectures, readings, and discussions.	
Extrinsic goal orientation	If I can, I want to get better grades in my classes than most of the other students.	.419
	I want to do well in my classes because it is important to show my ability to my family, friends, employer, or others.	
	Getting a good grade in my classes is the most satisfying thing for me right now.	

	The most important thing for me right now is improving my overall grade point average, so my main concern in my classes is getting a good grade.	
Help seeking	I try to identify students in my classes whom I can ask for help if necessary.	.731
	Even if I have trouble learning the material in a class, I try to do the work on my own, without help from anyone. (REV)	
	When I can't understand the material in a course, I ask another student in the class for help.	
	I ask the instructor to clarify concepts I don't understand.	
Intrinsic goal orientation	The most satisfying thing for me in my courses is trying to understand the content as thoroughly as possible.	.665
	When I have the opportunity in my classes, I choose course assignments that I can learn from even if they don't guarantee a good grade.	
	I prefer course material that arouses my curiosity, even if it is difficult to learn.	
	I prefer course material that really challenges me so I can learn new things.	
Metacognition	I try to change the way I study in order to fit the course requirements and the instructor's teaching style.	.747
	I often find that I have been reading for this class but don't know what it was all about.	
	I ask myself questions to make sure I understand the material I have been studying in my classes.	
	When I become confused about something I'm reading for a class, I	

	go back and try to figure it out.	
	If course readings are difficult to understand, I change the way I read the material.	
	During class time I often miss important points because I'm thinking of other things. (REV)	
	I try to think through a topic and decide what I am supposed to learn from it rather than just reading it over when studying for a course.	
	When I study for a class, I set goals for myself in order to direct my activities in each study period.	
	When studying for a course I try to determine which concepts I don't understand well.	
	If I get confused taking notes in class, I make sure I sort it out afterwards.	
	When reading for a course, I make up questions to help focus my reading.	
	Before I study new course material thoroughly, I often skim it to see how it is organized.	
Motivation (orig.)	If my teacher doesn't assign something I probably won't do it, even if it might help me learn the material. (REV)	.707
	I do assignments the teacher says are optional even if they don't earn extra credit.	
Organization	When I study for a course, I go through the readings and my class notes and try to find the most important ideas.	.508
	When I study the readings in a course, I outline the material to help me organize my thoughts.	

	I make simple charts, diagrams, or tables to help me organize course material.	
	When I study for a course, I go over my class notes and make an outline of important concepts.	

Peer learning	I try to work with other students from my classes to complete the course assignments.	.762
	When studying for a course, I often set aside time to discuss course material with a group of students from the class.	
	When studying for a course, I often try to explain the material to a classmate or friend.	
Self-testing (orig.)	I quiz myself with practice questions in textbooks to help me memorize class material.	.538
	I quiz myself by using flashcards to help me memorize class material.	
Rehearsal	I memorize key words to remind me of important concepts in my classes.	.544
	When studying for a course, I read my class notes and the course readings over and over again.	
	When I study for a class, I practice saying the material to myself over and over.	
	I make lists of important items for my courses and memorize the lists.	
Self-efficacy	I'm confident I can learn the basic concepts taught in my classes.	.903

	I believe I will receive excellent grades in my classes.	
	I expect to do well in my classes.	
	I'm confident I can understand the most complex material presented by the instructor in my courses.	
	I'm confident I can do an excellent job on the assignments and tests in my courses.	
	I'm certain I can master the skills being taught in my classes.	

	I'm certain I can understand the most difficult material presented in the readers for a course.	
	Considering the difficulty of my courses, the teacher, and my skills, I think I do well in my classes.	
Task value	I like the subject matter of my courses.	.728
	I think I will be able to use what I learn in one course in other courses.	
	It is important for me to learn the course material in my classes.	
	I think the course material in my classes is useful for me to learn.	
	I am very interested in the content area of my courses.	
	Understanding the subject matter of my courses is very important to me.	
Tech embrace (orig.)	I am always using the internet.	.858
	I prefer to do tasks by hand rather on the computer. (REV)	

	I am excited by new technologies.	
	Whenever possible, I like to own the newest technology.	
	I avoid using technologies that seem complicated.	
	If my family member has problems with their computer, I can help them.	
	I consider myself good with computers.	
	I enjoy learning ways that technology can make tasks more efficient.	
	I use tools on educational websites to help me study.	
Test anxiety	I have an uneasy, upset feeling when I take an exam.	.755
	When I take a test I think about how poorly I am doing compared with other students.	
	When I take a test I think about items on other parts of the test I can't answer.	
	I feel my heart beating fast when I take an exam.	
	When I take tests I think of the consequences of failing.	
Time management	I make sure that I keep up with the weekly readings and assignments for my courses.	.785
	I find it hard to stick to a study schedule. (REV)	
	I often find that I don't spend very much time on my courses because of other activities. (REV)	
	I make good use of my study time for my courses.	
	I rarely find time to review my notes or readings before an exam. (REV)	
	I attend my classes regularly.	
	I usually study in a place where I can concentrate on my course	

	work.	
	I have a regular place set aside for studying.	

Appendix C:

Study Strategy Survey, Version 1

Extra Credit: Study Strategy Survey.
Your responses will remain anonymous, will not impact your grade, and will not be shared without your explicit consent, which will be asked for in the coming days (and which you may decline without losing credit). In the meantime, please answer these questions honestly.

Listed here are several study strategies you might have used when preparing for this test. Refer to them by letter when answering the following questions:

A. Rereading notes or textbook
B. Do practice problems, self-tests, or flashcards
C. Re-write notes
D. Study with a group of students
E. Use mnemonics (acronyms, rhymes, etc.)
F. Make outlines or review sheets
G. Highlight (in notes or book)
H. Think of real life examples

1. Think of the strategies you used to prepare for **this test**. Rank the top 5 strategies in order of how often you used them, with 1 as most often. Only use numbers 1-5 once.

A	B	C	D	E	F	G	H
—	—	—	—	—	—	—	—

2. Out of all the study strategies (A-H), which do you think lead to the best learning outcomes **for the average student**? Rank the top 5 strategies, with 1 as best. Only use numbers 1-5 once.

A	B	C	D	E	F	G	H
—	—	—	—	—	—	—	—

3. Out of all the study strategies (A-H), which do you think lead to the best learning outcomes **for you**? Rank the top 5 strategies, with 1 as best. Only use numbers 1-5 once.

A	B	C	D	E	F	G	H
—	—	—	—	—	—	—	—

4. How beneficial do you believe the *Rereading notes or textbook* strategy is **for your** memory? Circle the number that best matches your opinion.

Not at all effective						Very effective
1	2	3	4	5	6	7

5. How beneficial do you believe the *Do practice problems, self-tests, or flashcards* strategy is **for your** memory? Circle the number that best matches your opinion.

Not at all effective						Very effective
1	2	3	4	5	6	7

Study Strategy Survey, Version 2

Extra Credit: Study Strategy Survey (__ points)
The following survey is optional. Your responses will remain anonymous and will not be shared without your consent, which you will be asked for in the coming days (and you may decline without losing credit). **Please answer honestly.**

1. Listed here are study strategies you might have used when preparing for this test. Think of the strategies that you used to prepare for THIS TEST. Rank the top 5 in order of how often you used them. (*Bubble in ONE PER ROW, and DO NOT choose any strategy more than once*)

	Rereading notes or textbook	Do practice problems, self-tests, or flashcards	Re-write notes	Study with a group of students	Use mnemonics (acronyms, rhymes, etc.)	Make outlines or review sheets	Highlight (in notes or book)	Think of real life examples
Most often used	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
2nd most often used	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
3rd most often used	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
4th most often used	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
5th most often used	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ

2. Out of all the study strategies listed here, which do you think lead to the best learning outcomes for the AVERAGE STUDENT? Rank the top 5 strategies. (*Bubble in ONE PER ROW, and DO NOT choose any strategy more than once*)

	Rereading notes or textbook	Do practice problems, self-tests, or flashcards	Re-write notes	Study with a group of students	Use mnemonics (acronyms, rhymes, etc.)	Make outlines or review sheets	Highlight (in notes or book)	Think of real life examples
Very best strategy	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
2nd best strategy	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
3rd best strategy	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
4th best strategy	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
5th best strategy	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ

3. Out of all the study strategies listed here, which do you think lead to the best learning outcomes for YOU? Rank the top 5 strategies. (*Bubble in ONE PER ROW, and DO NOT choose any strategy more than once*)

	Rereading notes or textbook	Do practice problems, self-tests, or flashcards	Re-write notes	Study with a group of students	Use mnemonics (acronyms, rhymes, etc.)	Make outlines or review sheets	Highlight (in notes or book)	Think of real life examples
Very best strategy	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
2nd best strategy	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
3rd best strategy	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
4th best strategy	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ
5th best strategy	Ⓐ	Ⓑ	Ⓒ	Ⓓ	Ⓔ	Ⓕ	Ⓖ	Ⓗ

4. How beneficial do you believe the *rereading notes or textbook* strategy is FOR YOUR memory? (*On the following scale, bubble in the ONE number that best matches your opinion*)

Not at all effective ① ② ③ ④ ⑤ ⑥ ⑦ Very effective

5. How beneficial do you believe the *do practice problems, self-tests, or flashcards* strategy is FOR YOUR memory? (*On the following scale, bubble in the ONE number that best matches your opinion*)

Not at all effective ① ② ③ ④ ⑤ ⑥ ⑦ Very effective

⓪	⓪	⓪	⓪	⓪	⓪	⓪	⓪	⓪
①	①	①	①	①	①	①	①	①
②	②	②	②	②	②	②	②	②
③	③	③	③	③	③	③	③	③
④	④	④	④	④	④	④	④	④
⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤	⑤
⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥	⑥
⑦	⑦	⑦	⑦	⑦	⑦	⑦	⑦	⑦
⑧	⑧	⑧	⑧	⑧	⑧	⑧	⑧	⑧
⑨	⑨	⑨	⑨	⑨	⑨	⑨	⑨	⑨

To earn credit, please BUBBLE in your EXACT Student ID number.

You will be asked to complete a similar extra credit survey on the next exam.

A corresponding extra credit online assignment will also be announced in the coming days.

PLEASE COMPLETE, DETACH, AND TURN IN DURING THE EXAM PERIOD.