Journal of Applied Research in Memory and Cognition xxx (2020) xxx–xxx

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off

Ayanna K. Thomas*
Tufts University, United States

Amy M. Smith
Quinnipiac University, United States

Kanika Kamal
Harvard University School of Medicine, United States

Leamarie T. Gordon
Assumption College, United States

This study examined whether frequent testing would promote long-term retention of college-level course material. Students in a college course engaged in three different types of interval practice over the course of a 13-week semester: quizzes, quizzes with feedback, and study. We examined the impact of type of interval practice on performance on unit exams. Six exams were given that consisted of multiple choice (MC) questions presented during earlier practice, new highly related MC questions, and new highly related short answer (SA) questions. The variation in type of unit exam questions allowed for the examination of interval quiz-related transfer to related MC and related SA questions. Further, half of the unit exams were taken individually and half were taken collaboratively. This manipulation allowed us to examine post-collaborative facilitation. Results suggest that interval quizzing resulted in beneficial transfer effects to highly related MC and SA questions and post collaborative facilitation.

General Audience Summary

Traditional undergraduate education in the United States has received increased criticism. The structure of these courses has been one of the primary focuses of criticism, because that structure has remained relatively stagnant for decades, taking the form of several weeks of lecture, followed by examination. This structure has been criticized for being outdated and ineffective in promoting long-term retention of material. Alternative approaches whereby frequent testing is used to promote learning and retention have been applied in the classroom, and results from these studies suggest that educators should consider employing this approach. The present study examined whether frequent testing would promote long-term retention of college-level course

Author Note.

The research reported in this paper was conducted with the support of a Faculty Research Award from Tufts University.

* Correspondence concerning this article should be addressed to Ayanna K. Thomas, Tufts University, Medford, MA 02155, United States. Contact: ayanna.thomas@tufts.edu.

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

material. Further, the present study examined three elements of test-related transfer to assess the breadth of testing utility. Students in a college-level Cognitive Psychology course engaged in retrieval or study practice during individual class meetings. Frequent testing facilitated memory and comprehension and promoted learning of new related concepts.

Keywords: Retrieval practice, Transfer effects, Classroom learning

In the present paper, we report a field-based, semester-long study that used retrieval practice to improve students' learning of college-level course material. Of the retrieval practice studies conducted in a live classroom, only a small number have examined testing effects over the course of an entire semester (e.g., Bangert-Drowns, Kulik, & Kulik, 1991; Carpenter, 2009; Foss & Pirozzolo, 2017; Gaynor & Millham, 1976; Keys, 1934; Leeming, 2002; McDaniel, Wildman, & Anderson, 2012; Pennebaker, Gosling, & Ferrell, 2013). Fewer still have investigated the effect in the context of college-level instruction (Mayer et al., 2009; McDaniel et al., 2012; Pan et al., 2019). Although these studies have provided important evidence that retrieval practice can be both employed in a semester-long course and used for improving mastery of college-level material, the evidence is limited to these few studies.

Therefore, the authors of this paper took the challenging step to redesign a previously taught college-level course to determine whether the inclusion of retrieval practice would be an effective and appropriate use of valuable course time. Specifically, we investigated whether frequent, in-class, low-stakes quizzing would promote mastery of complex theoretical constructs inherent to college-level courses. Traditional paper-and-pencil quizzes were given. In addition to our general goal of examining the feasibility of retrieval practice in the classroom, we sought to determine whether quizzing would confer benefits beyond enhancing memory of the tested material, and transfer across three specific and educationally relevant domains. Specifically, we examined (a) whether retrieval practice would confer benefits to later testing of related concepts, (b) whether retrieval practice would confer benefits when test format changed, and (c) whether student collaboration during retrieval practice would confer benefits to individual testing.

Transfer to Related Concepts

everal studies suggest that retrieval practice does confer transfer effects to related information. For example, in a laboratory-based study, Butler (2010) demonstrated that retrieval practice led to better learning and retention of related concepts within the same and in different knowledge domains as compared to study practice. However, transfer effects are not always found. In a laboratory experiment using a general biology text, Wooldridge, Bugg, McDaniel, and Liu (2014) found that retrieval practice did not confer an advantage over study practice when participants were assessed on topically related, but not identical information. Similarly, when test-bank questions associated with published course material were used, researchers found no benefit of prior retrieval practice over highlighting while studying on a final exam that tested new questions (Nguyen & McDaniel, 2015). However, when quiz items and final test items targeted the same concept, Nguyen and McDaniel reported a transfer effect.

Whereas performance on new and different concepts may not benefit from retrieval practice, a recent study by Foss and Pirozzolo (2017) found that retrieval practice resulted in better performance on final test questions when those final questions were related to practice items as compared to unrelated. The importance of strength of relationship between tested and untested concepts as a moderator for the impact of retrieval practice on performance was supported by a recent meta-analysis by Pan and Rickard (2018). Our goal was to further explore the value of retrieval practice on transfer to related concepts using materials that are typically used in classroom settings. Like Wooldridge et al. (2014) we used materials extracted from normed test-banks in which items were related at a general topic level. However, unlike Wooldridge et al., our related question pairs were also established such that highly related information was needed to answer both questions in the pair and developed from adjacent material in the text. In this way, our question pairs were more closely aligned with those used by Butler (2010) in that target material that was initially quizzed was the same material used to create unit exam questions.

Transfer Across Test Format

Recent research suggests that the value of retrieval practice may be apparent even when test format between initial and later testing changes. For example, researchers have found that middle school students were more likely to correctly answer chapter (Roediger, Agarwal, McDaniel, & McDermott, 2011) and unit exam short-answer (SA) questions (McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013; McDermott, Agarwal, D'Antonio, Roediger, & McDaniel, 2014) if information covered in those questions appeared on earlier multiple-choice (MC) quizzes, as compared to information that had not been quizzed earlier. Similarly, Smith and Karpicke (2014) found that practicing retrieval improved performance regardless of differences in test format. They further demonstrated that format mattered only when initial retrieval success for different formats was held constant. That is, initial retrieval success influenced later performance regardless of test format differences.

One of our goals was to provide additional evidence that superficial test format consistencies may have little impact on retrieval practice success. In the present study, students were exposed to inference and application MC questions during the retrieval practice phase and were later presented SA questions that tested highly related concepts. Although we did not employ Latent Semantic Analysis to quantify this relationship,

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

we developed questions that tested concepts that were presented in adjacent text.

Collaborative to Individual Testing

A last, but arguably as important, transfer effect under investigation in the present study was transfer between collaborative retrieval and individual test taking. Laboratory research suggests that engaging in collaborative testing during learning positively influences performance on later individual exams (e.g., Blumen & Rajaram, 2009; Blumen & Stern, 2011; Congleton & Rajaram, 2011; Wissman & Rawson, 2016). For example, collaborative testing in small groups conferred post-collaborative benefits on individual memory for word lists (Blumen & Stern, 2011; Wissman & Rawson, 2016), and even for more complex scenarios (e.g., Meade, Nokes, & Morrow, 2009). In a recent laboratory study, Wissman and Rawson (2018) found that collaborative testing had the greatest post-collaborative benefit when feedback was eliminated, suggesting that situations in which immediate feedback is given will reduce the benefits of collaborative testing on later individual testing. Although research suggests that collaborative test taking is preferred by students (Wissman & Rawson, 2016) and confers benefits to individual memory, the impact of collaboration on subsequent individual performance is absent from field-based research. The present study fills this important gap in the literature.

The Present Experiment

The present study investigated whether MC quizzes would enhance learning of Cognitive Psychology course content. The number of studies investigating the value of retrieval practice in the college classroom over an entire course remains understandably small. In the present study, we added to this body of research by assessing learning of course content using six in-class unit exams, and a comprehensive final exam. The course provided an in-depth exploration of classic and current issues in human cognition, examining behavioral, neuropsychological, and neuroscientific approaches to data and theory. Topics included experimental and neuropsychological methodologies in cognitive science, sensation, perception, attention, working memory, encoding and retrieval processes, implicit memory and multiple memory systems, categorization, decision making, developmental changes in cognition, and language acquisition and comprehension.

The course was structured to examine the effects of interval study or quizzing on learning. During each class meeting, the instructor gave two 20 min lectures. Immediately after each of the two short lectures, students were given a study sheet to review or a short MC test. After a series of quizzes and lectures, students were given a unit exam. The unit exam covered information from the previous series of lectures. A course-cumulative final exam was given at the end of the 13-week long semester.

Method

Participants

The participants were 54 students enrolled in Cognitive Psychology at Tufts University during the Fall 2014 semester ($M_{age} = 18.26$, $SD = 2.67$, range = 17–23). The data for nine students were dropped from analysis. One student had a documented learning disability that prevented participation, one student suffered a head trauma during the semester, and two students did not provide informed consent. The remaining students were dropped because they had missed at least one class meeting that included interval practice which included quizzing, quizzing with feedback, or study depending on group assignment. Thus, all analyses were conducted on the remaining 45 students who all provided informed consent and attended all class meetings that had interval practice.

Design and Procedure

The experiment employed a 3 (interval practice condition: study, quiz, quiz with feedback) × 2 (cumulative exam type: collaborative, individual) between-subjects factorial design. Additionally, student participants were assigned a participation number for the entire semester and divided into three separate groups. Group division was necessary in order to counterbalance participation in each interval practice condition. All participants were exposed to interval study practice, interval quiz practice, and interval quizzes with feedback. However, as Figure 1 illustrates, exposure to a specific practice condition alternated across three experimental blocks which correspond to time across the semester.

The experiment was conducted in the context of the undergraduate Cognitive Psychology course at Tufts University. The experiment lasted for the duration of the Fall 2014 academic semester, which began on September 2, 2014 and ended on December 18, 2014, which consisted of 13 weeks of class meetings. The class met from 6:00–7:15pm on Mondays and Wednesdays, with the exception of university holidays that occurred on October 13, November 11, and November 26.

On the first day of class, all students were given the course syllabus and informed consent. A modified version of each of these forms can be found in sclicker quizzing on jargon terms enhances definition-focused but upplementary materials. Students were informed that the course would proceed as specified in the syllabus, and informed consent was to be given if students agreed to have their exam performance used as data after the semester ended. To reduce the influence of coercion, informed consent forms were collected by a student volunteer at the end of the first class meeting. Those forms were placed in a sealed envelope and given to the Psychology Department administrator. The envelope was given to the first author and course instructor after final grades associated with the course were posted. Thus, the course instructor and TA were not aware of consent until the course ended. An alternative to the testing procedure was offered to students who were granted special accommodations for testing. No student chose the alternative procedure.

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

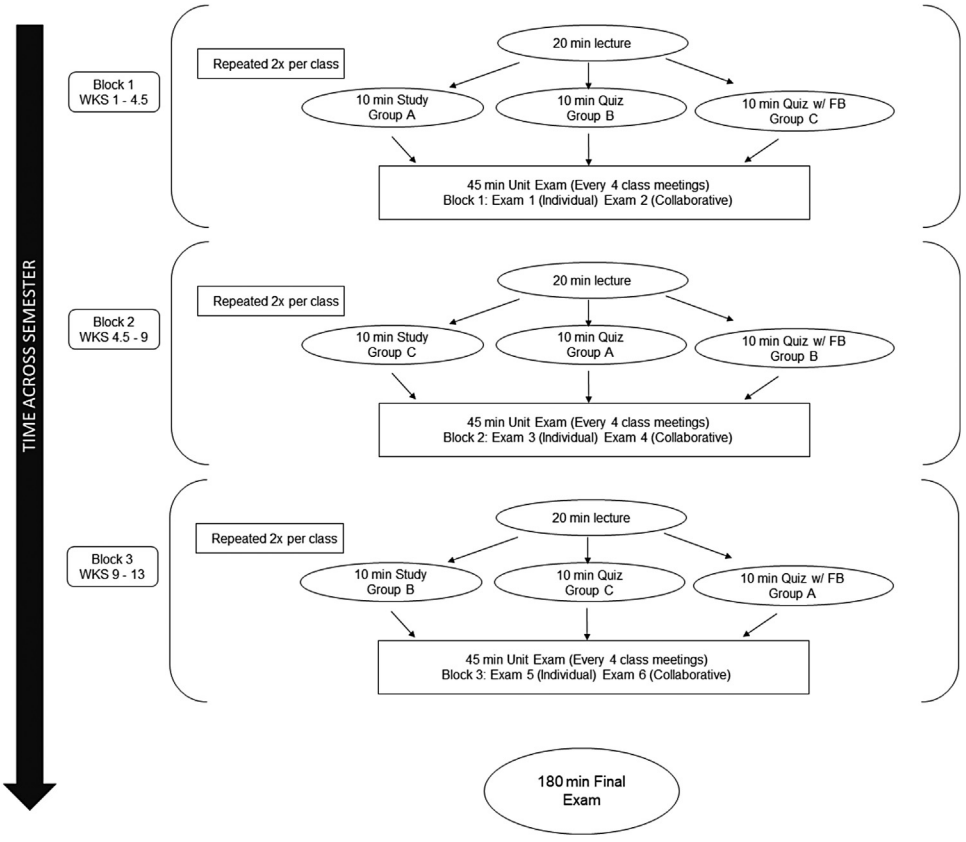

Figure 1. General methodology and design schematic.

All students were assigned a participant identification number on the first day of class. As opposed to names, students were required to use participation identification numbers on all exam materials. Participants were then randomly assigned to one of three groups. Group assignment was necessary because each group was assigned to different interval practice conditions during different weeks of the semester. For example, during weeks one through four, Group A was assigned to the study practice group. During weeks five through eight, Group A was assigned to quiz practice group. During weeks nine through 12, Group A was assigned to quiz with feedback group. Similarly, during weeks one through four, Group B was assigned to the quiz group and Group C was assigned to quiz with feedback. During weeks five through eight, Group B was assigned to quiz with feedback and Group C was assigned to the study group. During weeks nine through 12, Group B was assigned to the study group and Group C was the quiz group. Weeks one through four will be referred to as Block 1. Weeks five through eight will be referred to as Block 2. Weeks nine through 12 will be referred to as Block 3. Students were instructed to sit in the same seats throughout the semester and to sit with their assigned groups. Thus, within a given block, between-participants statistical comparisons can be made. Although within-participants analysis could also be conducted across the semester, we opted to not conduct these analyses because they are confounded by time during the semester and course material.

Each class meeting took one of two forms: (a) on *quiz* days, students listened to a 20 min lecture, took a 10 min quiz or engaged in study practice, listened to another 20 min lecture, and lastly took another 10 min quiz or engaged in study practice; (b) on unit *exam* days (not including the final exam), students listened to a 20 min lecture and then completed a 45-min exam. Unit exam days occurred after every four interval practice days.

At the beginning of a class meeting, the instructor would begin the lecture. After the first lecture sequence, there was a brief period used to answer questions or clarify points made in the lecture, and then the first quiz or study sheet for the class meeting was distributed. Students were instructed to not look at the distributed material until told to do so. When distribution was completed, students were instructed to begin review or quizzing. Students were instructed to continue to review the document until told to stop. After 10 minutes, students were told to turn over their sheets and these sheets were collected. The instructor then began the second 20 min lecture that was followed by a second interval practice session.

After each class meeting, study sheets were posted on Trunk, the Tufts University virtual course management system (https://trunk.tufts.edu), where all students could digitally access them for the remainder of the semester. The TA graded all quizzes immediately after each class meeting and quiz grades were released to students via Trunk. At each subsequent class meeting, graded quizzes were returned only to students in the *quiz with feedback* condition. Therefore, students in the feedback condition received information about which questions they got correct or incorrect in the next class meeting. All other students had to infer individual question accuracy. Feedback consisted of

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

marking incorrect responses as wrong and providing the correct answer. As this form of feedback had no measurable impact on performance, it will not be discussed in any further detail. The exam conditions were counterbalanced within subjects such that each student engaged in studying, quizzing, and quizzing with feedback each for one-third of the semester. This design was employed to ensure that all students had access to the same elements of instruction. In addition to the course syllabus, students were provided with a class meeting schedule that indicated how each group would practice during a given class meeting. The schedule also indicated when students would be taking individual and collaborative unit exams.

On each of the six unit exam days, the instructor gave a 20-min lecture and then distributed the 48-item unit exam. Content from the preceding 20-min lecture was tested on the subsequent unit exam. Students completed three of the unit exams individually and completed the other three exams in randomly assigned groups. Students were assigned to new groups for each collaborative exam, such that they took each collaborative exam with a unique group of people. Groups consisted of three to five students. For collaborative exam taking, students were instructed that they could discuss all questions with members of their group; however, each student was required to turn in a separate exam. We alternated whether unit exams were taken individually or collaboratively, starting with the first exam which was completed individually. As with the quizzes, the unit exams were completed on paper and distributed by the TA. Students were told when to begin, were required to review the exam until the 45 allotted minutes for testing had elapsed, and the exams were collected by the TA.

The final exam was administered via Qualtrics (2018) on December 15, 2014, during Tufts University's scheduled finals week. Students completed the exam either in a computer lab or in a classroom using their own personal laptops. The TA monitored students who took the exam in the computer lab and the instructor monitored students who took the exam in the classroom. Students took between 90 and 150 min to complete the final exam. This was the only exam that was taken via computer.

Materials

Quizzes. Over the course of the semester, a total of 34 quizzes were created and implemented. Quizzes consisted of 10 MC questions chosen from the test bank that accompanied the instructor copy of Reisberg's (2013) *Cognition* textbook. The course instructor selected the questions for each quiz. Quiz questions were related to the lecture material that had been covered during the lecture that preceded each quiz. Quizzes were designed to take 10 min to complete. Questions on quizzes were conceptual in nature and included a mixture of application and inference questions. Different types of application questions included *analysis or evaluation* (i.e., interpreting new data or an example in the context of prior learning), *comparison or contrast* and *prediction* (i.e., determining how a system is affected by a new situation). Different types of inference questions include *bridging inferences* (i.e., integrating multiple pieces of information that were presented separately) and *conceptual inferences* (i.e., uncovering an underlying or overall principle). The designation of question type was informed by Barnett and Ceci (2002) and evaluated in the context of Pan and Rickard (2018). In addition, the test bank provided information about question type, and all questions were independently categorized as application or inference questions by the first author. Inference and application questions were grouped together for statistical analysis. A sample set of quiz questions can be found in Appendix A. Questions were presented to all students in a fixed order, and students completed these quizzes via paper and pen.

Study sheets. Content-matched study sheets were also presented over the course of the semester. As with quizzes, 34 study sheets were created. Study sheets were matched in content to accompanying quizzes. To create the study sheets, quiz questions were modified into statements that contained the correct answer. The incorrect, alternative answer choices were not included. Thus, students who studied were not exposed to the alternative choices that were presented to those in the two test conditions. A sample sheet can be found in Appendix B. Study items were presented to students in a fixed order.

Unit exams. Throughout the semester, six unit exams were created and implemented. Each unit exam featured 48 questions: 15 MC questions that were repeated from quizzes given in the preceding five lectures, and 15 new MC questions that also tested knowledge of the previous five lectures. An additional 18 SA questions were used to assess test-format transfer in the context of related-concept transfer. All repeated and new multiple-choice questions were presented in a fixed order at the beginning of each cumulative exam, followed by all SA questions. Students were given printed exams and completed using pens. Sample questions from unit exams can be found in Appendices C and D.

The 15 MC questions designed to assess related concept transfer did assess information that was initially studied in the context of the chapter readings and covered in lecture. However, these questions had not been asked during interval practice. Further, study sheets did not provide answers to these questions. Related questions were designed to test concepts from the text book presented in adjacent portions of the text materials (e.g., same subsection of a chapter). In addition, the relationship of paired questions was determined by two independent raters. Only questions for which the two raters agreed were considered related.

Final exam. The final exam consisted 104 multiple choice questions followed by six essay questions. The 104 multiple choice questions consisted of eight questions from each of the 13 chapters in the Reisberg (2013) textbook. The eight questions chosen from each chapter consisted of one *quiz-only* question that was repeated from a previous quiz, four *unit-only* questions that were repeated from a previous unit exam, one question that had been presented on a previous quiz and later repeated on a unit exam, and two new questions. The final exam was administered via Qualtrics (2018). Questions were presented eight at a time, blocked by chapter. The eight questions within each chapter were presented in a fixed order. The order of the questions was as follows: quiz-only, new, unit-only, new, both. The 13 chapter blocks were presented in random order. Following the MC

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

questions, participants were presented with each essay question, and order of presentation was randomized. Sample final exam essay questions can be found in Appendix E.

Course syllabus. The course syllabus mapped out the structure of each class meeting, when quizzes and exams would take place, and how final grades would be calculated. A modified version of the course syllabus can be found in the online supplementary materials. The course syllabus provided students an overview of the course and discussed how students should prepare for each class meeting. Importantly, students were encouraged to come to each class meeting having read the required material in advance of the meeting. Students were informed that at two points during a given class meeting they would have the opportunity to study or take a quiz on concepts that had been covered in the previous reading and lecture. Students were informed that they would have 10 minutes for each study/quiz break. Students were further informed that each quiz was worth 1.5% of their final grades. Although 23 quizzes were given, we counted 20 toward a student's final grade. Students were also informed that each unit was worth 6% of their final grade. Attendance also factored into a student's final grade to encourage consistent participation across the semester.

Exam scoring. All quizzes and exams were scored by both the TA and course instructor. We employed scantrons for all MC quizzes and unit exams, with the exception of the final exam. SA questions were constructed to have one correct answer and could be completed with one or two words. SA scores were either correct or incorrect. Partial credit was not given. The final exam was given via computer and was scored by the instructor. The MC questions were scored via a computer program that compared responses to an answer key. The essays were scored using a rubric designed by the course instructor. Since these essays were not analyzed for this paper, we have chosen not to include further detail regarding scoring these answers.

Results

Transfer Effects

In this first section, we examined the transfer effects as described in the Introduction section (i.e., transfer to new but related concepts for MC questions, transfer across question format and related concepts, transfer between collaborative and individual exam taking). First, we examined whether quizzes would result in transfer benefits to related concepts tested in the same format. Second, we examined whether quizzes would result in improved learning when the format of later exam questions changed. Third, we examined whether collaborative exam taking would result in better performance on identical questions presented on the final exam as compared to individual exam taking. To preview these results, we found that quizzing was associated with better performance on later-tested, related concepts as compared to studying (Figure 2). We found that quizzing resulted in better performance than studying when exam format changed (see Figure 3). Finally, we found evidence for post-collaborative benefits. That is, when students took unit exams in groups, they performed better on identical MC questions as compared to students who took individual unit exams.

Transfer to Related Concepts on Individual Unit Exams. To examine whether quizzes facilitated performance on unit exams, quiz and exam questions were developed in tandem to assess highly related concepts. For all related questions presented on unit exams, a quiz question had been previously presented that covered the same conceptual material. These question pairs assessed information from the same subsection of a chapter and were always adjacent concepts from the text. For example, two questions that assessed some aspect of Baddeley's working memory model were considered related. Further, transfer to related concepts was examined by analyzing performance on the three unit exams taken individually (unit exams 1, 3, 5). Finally, because the three exams were taken at different points in the semester, we conducted separate analyses for each individual unit exam. This resulted in three one-way, between-participants ANOVAs that compared performance on related MC questions presented on three separate unit exams. For unit exam 1, we found a significant effect of practice, $F(2, 42) = 5.36$, $p < .001$, $\eta_p^2 = .20$. Pairwise comparisons, using a Bonferroni correction found that this main effect was driven by the significantly better performance of participants who had taken a quiz ($M = .83$) compared to participants who studied ($M = .71$), $t(28) = 3.57$, $p < .001$, $d = 1.26$. No other comparisons were significant when examined using corrected pairwise tests. The pattern of practice condition differences can be viewed in Figure 2.

For unit exam 3, we found a significant effect of practice condition, $F(2, 42) = 7.63$, $p < .001$, $\eta_p^2 = .27$. Pairwise comparisons using a Bonferroni correction found that participants who took quizzes ($M = .79$) performed better on related unit exam questions than participants who engaged in study ($M = .65$), $t(28) = 3.98$, $p < .001$, $d = 1.38$. We also found that quizzes with feedback ($M = .75$) led to statistically significant improvement on unit exam related questions as compared to study, $t(28) = 2.96$, $p = .006$, $d = .95$. The difference between quiz and quiz with feedback was not significant. Finally, for unit exam 5, we again found a main effect of practice condition, $F(2, 42) = 5.15$, $p = .01$, $\eta_p^2 = .20$. As is evident in Figure 2, the same pattern of results was found for unit exam 5 as found in the previous individual until exams. Further, pairwise comparisons using a Bonferroni correction found a significant difference between participants who took quizzes ($M = .81$) and those that engaged in study ($M = .67$), $t(28) = 3.52$, $p < .001$, $d = 1.27$.

Transfer Across Test Format on Individual Unit Exams. We were interested in whether MC quizzes during learning would result in better SA test performance on unit exams than study practice. SA questions were created to test concepts previously tested on MC quizzes. However, these questions could not be completed with the same answer as related MC questions presented during practice quizzes. As with the analysis for related concepts, MC and SA questions tested highly related concepts associated with adjacent material presented in chapters. As with the earlier analysis, we limited this analysis to individual unit exams and examined this aspect of transfer for each unit exam separately. For unit exam 1, we found a significant effect of practice condition, $F(2, 42) = 13.28$, $p < .001$, $\eta_p^2 = .96$. Pairwise comparisons, using a Bonferroni correction found that participants who were in the quiz condition ($M = .68$)

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

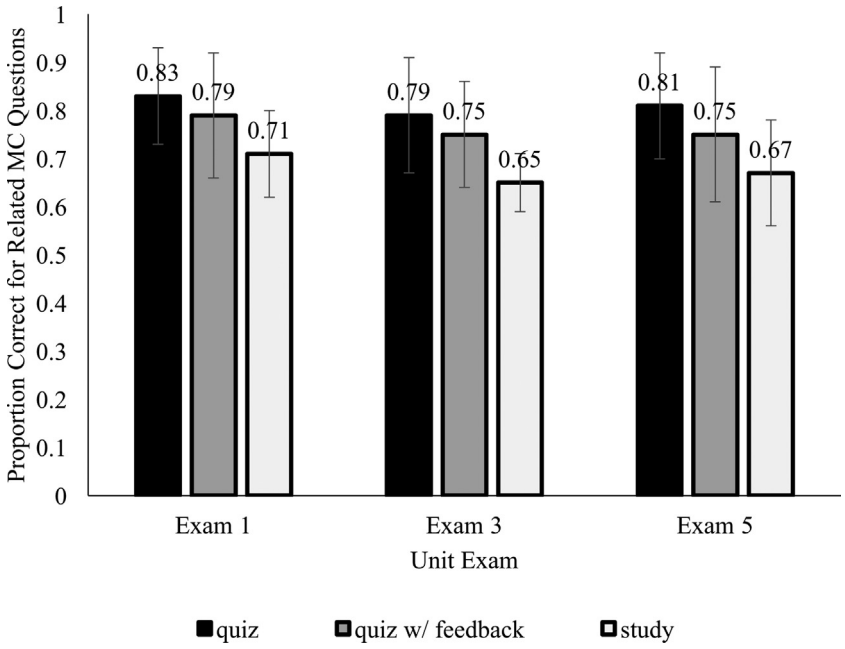

Figure 2. Transfer between related concepts presented during interval practice and tested as multiple-choice questions on unit exams. Transfer is plotted as a function of interval practice group for each unit exam taken individually. Means and SDs are represented.

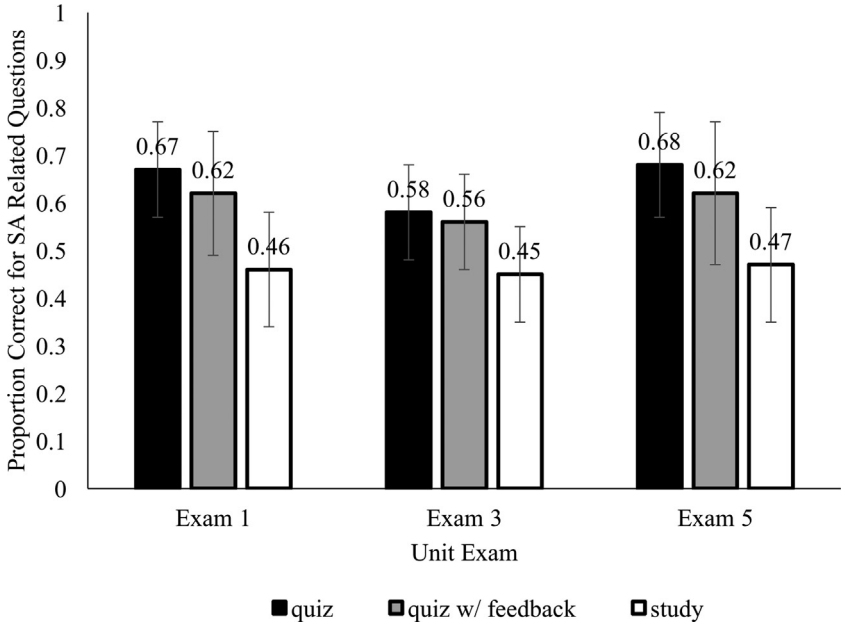

Figure 3. Transfer between related concepts presented during interval practice and tested as short answer questions on unit exams. Transfer is plotted as a function of interval practice group for each unit exam taken individually. Means and SDs are represented.

performed significantly better than participants in the study condition ($M = .46$), $t(28) = 5.36$, $p < .001$, $d = 2.17$. We also found an advantage for the quiz with feedback condition ($M = .62$) over the study condition, $t(28) = 3.46$, $p = .002$, $d = 1.28$. The difference between the two testing conditions did not reach statistical significance. The pattern of practice condition differences can be viewed in Figure 3.

For unit exam 3, we found a significant effect of practice condition, $F(2, 42) = 7.66$, $p < .001$, $\eta_p^2 = .27$. Pairwise comparisons using a Bonferroni correction found that participants who took quizzes ($M = .57$) performed better on related unit exam short answer questions than participants who engaged in study ($M = .45$), $t(28) = 3.61$, $p = .001$, $d = 1.30$. We also found that quizzes with feedback ($M = .56$) led to statistically significant improvement on unit exam related short answer questions as compared to study, $t(28) = 3.09$, $p = .005$, $d = 1.10$. The difference between quiz and quiz with feedback was not significant. Finally, for unit exam 5, we again found a main effect of practice condition, $F(2, 42) = 11.31$, $p = .001$, $\eta_p^2 = .35$. As is evident in Figure 3, the same pattern of results was found for unit exam

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

5 as found in the previous individual until exams. Participants who practiced with quizzes ($M = .68$) and quizzes with feedback ($M = .62$) did statistically better on related short answer questions than those who engaged in study practice ($M = .47$), $t(28) = 5.12$, $p < .001$, $d = 1.82$; $t(28) = 3.14$, $p = .004$, $d = 1.10$.

Post-Collaborative Benefits. A third motivation for this study was to investigate whether collaboration during unit exams would have benefits when tested on the same material at the end of the semester. In order to examine potential post-collaborative benefits, the six unit exams (three taken individually and three taken in collaboration) included a set of MC questions associated with each chapter in a unit that had not been previously presented. The reader should note that these MC questions are the same as discussed in an earlier section in the results that focused on transfer to related concepts. In the case of the analysis focused on post-collaborative benefits, we were interested in whether collaborating on these questions would influence performance on these *same* questions on a later final exam. Eight questions from unit exams 1 through 5 were selected for the final. Twelve questions from unit exam 6 were selected for the final. Additional questions from unit exam 6 were included on the final because this unit exam covered three chapters, whereas the previous unit exams covered two chapters. The reader should also be aware that because the manipulation of collaboration varied with material, there is a confound inherent in this design. Therefore, we opted to compare final test performance on questions taken from unit exams 1 (individual) and 2 (collaborative) in one analysis, exams 3 (individual) and 4 (collaborative) in a second analysis, and exams 5 (individual) and 6 (collaborative) in a third analysis. Although this approach does not eliminate the confound of material differences varying with collaboration, this approach does allow for an examination of post-collaborative benefits using different sets of materials.

A 2 (type of test: individual, collaborative) × 3 (type of practice: study, quiz, quiz with feedback) ANOVA was conducted on average proportion correct of questions presented on the final exam that had previously been presented on unit exam 1 (individual) and unit exam 2 (collaborative). We found a main effect of type of test, $F(1, 42) = 10.04$, $p = .003$, $\eta_p^2 = .19$. Participants who engaged in collaborative test taking ($M = .98$) did better on final exam questions that participants who engaged in individual test taking ($M = .88$). Though this effect was significant, we acknowledge that the materials did differ between the two types of test, and further acknowledge potential limitations in interpretation due to ceiling effect performance. Type of practice had no impact on performance on these questions, nor did it interact with type of test, F's < 1.

A similar analysis comparing performance on questions taken from exam 3 (individual $M = .87$) and 4 (collaborative $M = .96$) also found a main effect of type of test, $F(1, 42) = 4.77$, $p = .035$, $\eta_p^2 = .10$. As with the previous analysis, type of practice did not impact performance on these questions, nor did it interact with type of test, F's < 1. Finally, we found a main effect of type of test when exam 5 (individual $M = .87$) and exam 6 (collaborative $M = .96$) average question performance was compared, $F(2, 42) = 8.01$, $p = .007$, $\eta_p^2 = .19$. Again, type of initial practice had no bearing on this pattern. The pattern of results is striking. If participants engaged in collaborative testing, they performed better on questions presented during collaboration on a final test than if they engaged in individual testing. Although the comparisons were confounded by different materials between the two different types of test, the finding that the pattern holds regardless of which two sets of questions were compared suggests that the materials differences may not have substantially impacted these findings.

General Effects of Interval Practice

Although the primary focus of this research was on transfer effects, the design of the present study also allowed us to examine the standard testing effect.

Unit Exam Performance Between Interval Practice Conditions for all MC Questions. In order to examine the general effects of type of interval practice on unit exams, in this section we focused only on performance on MC questions on these exams as a function of practice condition assignment. Some of the questions on the unit exams were previously presented in the context of practice. Some of the questions were new, but related. Because participants rotated among the three practice conditions (study, quiz, quiz with feedback) across the semester, we were able to examine the effects of interval practice in a between-subjects design for each of the three counterbalancing blocks.

Block 1: Unit Exams 1 and 2. A 3 (practice: study, quiz, quiz with feedback) × 2 (unit exam: individual, collaborative) × 2 (type of question: repeated, related) ANOVA was performed on average unit exam performance associated with Block 1. Main effects for type of unit exam, $F(1, 41) = 75.51$, $p < .001$, $\eta_p^2 = .65$, question type, $F(1, 41) = 78.68$, $p < .001$, $\eta_p^2 = .66$, and practice, $F(2, 41) = 8.21$, $p < .001$, $\eta_p^2 = .29$, were found. Planned comparisons that employed a Bonferroni correction demonstrated that participants in the study condition ($M = .87$) performed less well than participants in the quiz condition ($M = .94$), $t(28) = 3.56$, $p < .001$, $d = 1.30$. Participants in the quiz with feedback condition ($M = .92$) also performed better than those in the study condition, $t(27) = 2.41$, $p = .023$, $d = 0.96$. The difference between the two quiz conditions was not significant. We also found a significant interaction between question type and unit exam, $F(1, 41) = 37.30$, $p < .001$, $\eta_p^2 = .47$. When collapsed across interval practice conditions, there was a greater improvement in performance on related questions between individual and collaborative testing as compared to the improvement in performance on repeated questions; however, caution should be taken when interpreting these results due to the ceiling effects apparent in the collaborative testing group. Means associated with Block 1 performance can be found in Table 1.

Block 2: Unit Exams 3 and 4. To preview, the pattern of results mirrors that found in Block 1. Main effects for unit exam, $F(1, 42) = 45.40$, $p < .001$, $\eta_p^2 = .52$, question type, $F(1, 42) = 45.97$, $p < .001$, $\eta_p^2 = .52$, and interval practice, $F(2, 42) = 6.67$, $p = .003$, $\eta_p^2 = .24$, were found. Planned comparisons that employed a Bonferroni correction also demonstrated that participants in the study condition ($M = .84$) performed less well than participants in the quiz condition ($M = .91$),

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

Table 1
Average Performance on Unit Exams for Repeated and Transfer Multiple Choice Questions as a Function of Testing Block and Interval Practice. Means and Standard Errors.

	Study	Interval quiz	Interval quiz with feedback
Block 1			
Quiz	–	.85 (.01)	.84 (.01)
Individual repeated	.91 (.02)	.92 (.03)	.98 (.01)
Collaborative repeated	.93 (.03)	1.00 (.00)	.99 (.01)
Individual related	.73 (.03)	.81 (.01)	.78 (.02)
Collaborative related	.90 (.02)	.97 (.02)	.97 (.02)
Block 2			
Quiz	–	.78 (.02)	.78 (.02)
Individual repeated	.86 (.02)	.95 (.04)	.93 (.03)
Collaborative repeated	.89 (.04)	1.00 (.00)	1.00 (.00)
Individual related	.70 (.03)	.79 (.04)	.72 (.03)
Collaborative related	.91 (.01)	.94 (.04)	.94 (.03)
Block 3			
Quiz	–	.81 (.01)	.81 (.02)
Individual repeated	.91 (.02)	.96 (.01)	.97 (.01)
Collaborative repeated	.99 (.01)	.96 (.01)	.99 (.02)
Individual related	.67 (.02)	.74 (.03)	.75 (.03)
Collaborative related	.82 (.02)	.96 (.01)	.91 (.01)

$t(28) = 3.14$, $p = .004$, $d = 1.15$. Participants in the quiz with feedback condition ($M = .90$) also performed better than those in the study condition, $t(28) = 3.10$, $p = .004$, $d = 1.14$. The difference between the two interval quiz conditions was not significant. Finally, we found a significant interaction between question type and unit exam, $F(1, 42) = 25.52$, $p < .001$, $\eta_p^2 = .39$. As in Block 1, when collapsed across interval practice condition, performance on related questions improved to a greater extent between individual and collaborative testing as compared to performance on repeated questions. Means associated with Block 2 performance can be found in Table 1.

Block 3: Unit Exams 5 and 6. Main effects for test, $F(1, 42) = 56.87$, $p < .001$, $\eta_p^2 = .58$, question type, $F(1, 42) = 247.89$, $p < .001$, $\eta_p^2 = .86$, and interval practice, $F(2, 42) = 5.40$, $p = .008$, $\eta_p^2 = .21$, were found. Participants in the quiz with feedback condition ($M = .90$) performed significantly better on the unit exams associated with Block 3 than participants in the study condition ($M = .85$), $t(28) = 3.25$, $p = .003$, $d = 1.20$. While not significant after the Bonferroni correction, there was a numeric difference in performance differences between the quiz and study conditions, $t(28) = 1.84$, $p = .08$. Finally, we found a significant interaction between question type, and unit exam, $F(1, 42) = 18.73$, $p < .001$, $\eta_p^2 = .31$. As in the previous blocks, when collapsed across interval practice condition, performance on related questions improved to a greater extent between individual and collaborative testing as compared to performance on repeated questions. Means associated with Block 3 performance can be found in Table 1.

General Discussion

Adding to the growing body of classroom-based experiments examining the value of repeated testing, the present study demonstrated that when in-class practice was retrieval-based, students performed better across different test formats and on new but related concepts, than when practice was study-based. We have also demonstrated that complex integrated college-level material does benefit from learning practices that incorporate retrieval, addressing criticism that retrieval practice is only useful for the recollection of previously retrieved information, or conceptually simple material (e.g., Gatto, 2011). In addition to enhancing knowledge for previously tested material, we have shown that interval quizzing resulted in better performance on new MC questions and SA questions that were conceptually related to previously tested concepts.

These findings suggest that retrieval practice may aid in the development of broader concept development. For example, the ability to synthesize information in order to answer conceptually related compare-and-contrast inference questions requires that students not only remember and understand aspects of two compared models, but also be able to apply their knowledge, organize concepts, and evaluate the two models (e.g., Bloom, 1956). When retrieval practice was used to support initial learning, students performed better on related inference and application questions that required a moderate level of conceptual understanding as compared to when they engaged in study practice. Though our results are indirect, they suggest that testing earlier in the process of learning helps to build a foundation of knowledge that students build upon and can sophisticatedly use weeks later.

While this study did not examine the underlying mechanism that may result in testing-related benefits to related concepts, research from members of this team and several others in the field suggest that taking a test potentiates learning of new and related concepts. In an eyewitness memory paradigm, Gordon and Thomas (2014) demonstrated that when participants took a structured cued recall test of a witnessed event, they learned subsequently presented related information to a better extent than participants who did not take the test. Just how testing potentiates new learning remains less well understood and uncovering the precise mechanism by which test-related transfer occurs is an important goal for future research (e.g., Pan & Rickard, 2018).

We also demonstrated that retrieval practice in a group resulted in post-collaborative benefits. To our knowledge, this is the first study to demonstrate post-collaborative benefits in a college classroom. This finding adds to the growing body of research demonstrating post-collaborative benefits. More importantly, this aspect of our study provides another useful technique that educators may employ to reduce the anxiety that often accompanies test-taking. Allowing students to collaborate in groups on unit exams resulted in better individual performance than when students were individually tested on the final exam.

Limitations

We would like to acknowledge that this investigation of the value of retrieval practice in a classroom setting has limitations. As with most field-based studies, isolating the cause of specific effects are challengingbecause the experimenters have little control of how students study outside of the classroom. We were

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

unable to control for external factors that may have affected student performance and engagement. For example, we had no control of how students engaged with the material outside of the classroom, the amount of time students could devote to their studies, whether students differentially prepared for individual as opposed to collaborative unit exams, and whether students were able to devote their full attention to all classroom practice.

The cognitive antidote principle suggests that when an additional cognitive requirement is introduced to what might be considered a tedious task, performance may improve because of changes in attentional focus (e.g., Kole, Healy, & Bourne, 2008). In a recent study Healy, Jones, Lalchandani, and Tack (2017) demonstrated that task engagement improved when quizzes were given shortly after learning as opposed to when quizzes were delayed. Similarly, Szpunar, Jing, and Schacter (2014) found that regular quizzing reduced incidences of self-reported mind wandering. Changes in learner engagement may also have downstream consequences on how students interact with the material outside of the classroom. Therefore, as opposed to considering the effects of retrieval practice in this study as direct, we may wish to consider the very strong possibility that retrieval practice resulted in indirect effects that included impacting engagement with the material inside and outside of the classroom.

These factors may have also contributed to why differences between quizzes and quizzes that included feedback were not found. That is, all students may have reviewed material after taking quizzes. In addition, although we provided quiz feedback at following class meetings, the time between class meetings varied between two and five days because the course met on Mondays and Wednesdays. Research suggests that the timing of feedback may play an important role in the effectiveness of feedback (e.g., Iwaki, Nara & Tanaka, 2017; Mullet, Butler, Verdin, von Borries, & Marsh, 2014; Sinha & Arnold, 2015). Finally, although we reported evidence for post-collaborative benefits, our conclusions were drawn from comparisons between individual and collaborative test taking on different sets of materials. The reader should be aware that materials difference may have influenced these results. That said, we did find post-collaborative facilitation across the three different Blocks in the semester.

Conclusion

The present study took advantage of college-level classroom-based learning in order to examine the benefits of interval quizzing on integrating new but related material. We measured this integration by examining the impact of interval quizzing on learning related concepts. Across the semester, we found that when students engaged in practice quizzing they performed better on related test questions than when they engaged in study practice. Therefore, we feel confident in concluding that quizzes resulted in transfer benefits to related concepts and across different test formats, and that testing in collaboration may foster post-collaborative benefits.

Conflicts of interest

The authors declare no conflict of interest.

Author Contributions

Ayanna Thomas and Leamarie Gordon developed the idea and hypotheses for the experiment. Ayanna Thomas developed the materials, methodology, and procedure. Ayanna Thomas and Amy Smith collected and scored raw data. Amy Smith and Kamika Kamal processed data and conducted preliminary analyses and literature reviews. Final analyses were conducted by Ayanna Thomas. Ayanna Thomas drafted the manuscript which was edited by Amy Smith and Leamarie Gordon.

Appendix A. Sample quiz practice

1. a. paralinguistic
   b. episodic
   c. semantic
   d. direct
   Answer: c
2. Which is NOT part of the Collins & Quillian model?
   a. nodes
   b. network
   c. bridging
   d. hierarchical organization
   Answer: c
3. "Defining feature" is most associated with _____.
   a. Bartlett's mental workbench
   b. connectionist models
   c. Collins and Quillian's hierarchical model
   d. Smith's feature list model
   Answer: d
4. In semantic memory tasks, response time is speeded up or judgments are made more easily when the concepts are closer together in semantic distance—that is, when they are more closely related. The effect is reversed when the comparison is false; that is, RT is longer for the comparison "a whale is a fish" vs. "a whale is a bird." This is an example of _____.
   a. superordinate effect
   b. semantic relatedness effect
   c. subordinate effect
   d. Hampton priming
   Answer: b
5. A person with anterograde amnesia would be expected to show _____ semantic priming effects, compared to normal controls.
   a. similar
   b. larger
   c. smaller
   d. adaptive
   Answer: a
6. In a priming experiment using lexical decision, what is the best "neutral" condition?
   a. truck–robin
   b. XXXX–dog
   c. France–Switzerland
   d. doctor–nurse
   Answer: b

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

7. Schemata aid in what aspect of memory?
   a. reconstructive processes
   b. reproductive recall
   c. analogical reasoning
   d. propositional coding
   Answer: a
8. Mental categories allow us to _____.
   a. predict the ways in which we should interact with new instances
   b. spend more time trying to figure out what things are
   c. find the needle in the haystack
   d. overcome our biases and prejudices derived from stereotypes
   Answer: a
9. In semantic categories, the degree to which items are viewed as typical, central members of a category; the central tendency of a category: _____.
   a. semantic activation
   b. typicality
   c. inheritance
   d. priming
   Answer: b
10. Rips (1975) reported an experiment in which subjects read a story about an island inhabited by only eight species of animals: sparrows, robins, eagles, hawks, ducks, geese, ostriches, and bats. The evidence indicated _____.
    a. support for prepositional theories of representation
    b. support for PDP models
    c. support for the dual-coding hypothesis
    d. evidence for prototype effects
    Answer: d

Appendix B. Sample study practice

1. The kind of memory that is thought to be largely similar across different people is semantic.
2. Bridging is NOT part of the Collins & Quillian model.
3. "Defining feature" is most associated with Smith's feature list model.
4. In semantic memory tasks, response time is speeded up or judgments are made more easily when the concepts are closer together in semantic distance—that is, when they are more closely related. The effect is reversed when the comparison is false; that is, RT is longer for the comparison "a whale is a fish" vs. "a whale is a bird." This is an example of semantic relatedness effects.
5. A person with anterograde amnesia would be expected to show similar semantic priming effects, compared to normal controls.
6. In a priming experiment using lexical decision, XXXX–dog is the best "neutral" condition.
7. Schemata aid in reconstructive processes of memory?
8. Mental categories allow us to predict the ways in which we should interact with new instances.
9. In semantic categories, the degree to which items are viewed as typical, central members of a category; the central tendency of a category: typicality.
10. Rips (1975) reported an experiment in which subjects read a story about an island inhabited by only eight species of animals: sparrows, robins, eagles, hawks, ducks, geese, ostriches, and bats. The evidence indicated evidence for prototype effects

Appendix C. Unit exam sample related questions

1. In a feature list model of semantic memory, the structure of semantic memory comes from _____.
   a. the nodes and links
   b. the structure of the lists and the retrieval processes
   c. the structure of the world
   d. the structure of our minds
   Answer: b
   Related to Question 3 in Appendix C
2. In testing their model, Collins & Quillian _____.
   a. used a lexical decision verification task
   b. were unable to account for serial exhaustive memory search functions
   c. demonstrated that concepts closer together in the network are responded to faster
   d. used both RT and accuracy measures
   Answer: b
   Related to Question 2 in Appendix C
3. Which of the following could be used as evidence AGAINST a "hierarchical" organization of semantic memory?
   a. serial position curve
   b. typicality effects
   c. hierarchical deconstruction
   d. cognitive economy
   Answer: b
   Related to Question 4 in Appendix C
4. Priming effects reveal what about semantic memory?
   a. when it was learned
   b. how it is structured
   c. when a connectionist network has transformed into a semantic network
   d. that there is functionally no end to semantic memory
   Answer: b
   Related to Question 5 in Appendix C
5. What nature of memory is best illustrated by the operation and influence of schemata during memory retrieval?
   a. forgetting
   b. compartmentalization
   c. learning
   d. reconstruction
   Answer: d
   Related to Question 7 in Appendix C

Appendix D. Unit exam sample short answer questions

1. Semantic memory captures _____ information. (GENERIC/GENERAL/ENCYCLOPEDIC)
   Related to Question 1 in Appendix C
2. Hintzman referred to semantic memory as _____ memory. (GENERIC)

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

Related to Question 1 in Appendix C

3. In a semantic network, concepts are represented by _____ and associations are represented by _____. (NODES; LINKS)

Related to Question 2 in Appendix C

4. In a priming task, the first stimulus is called the _____, and the next stimulus is called the _____. (PRIME; TARGET)

Related to Question 6 in Appendix C

5. What type of memory process does a schema or script aid in? (RECONSTRUCTIVE)

Related to Question 7 in Appendix C

6. The idealized average of all category members is called a(n) _____. (PROTOTYPE)

Related to Question 10 in Appendix C

Appendix E. Final exam essay questions

1. What is an advantage of feature list theories of semantic memory compared to semantic network models?
2. Describe how a priming task can be used to demonstrate both implicit and explicit processing.
3. How are schemata likely to change of the course of one's life?
4. In what ways are schemata and categories similar, and in what ways are they different?

Appendix F. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jarmac.2019.12.005.

References

Bangert-Drowns, R., Kulik, J. A., & Kulik, C. C. (1991). Effects of frequent classroom testing. *The Journal of Educational Research*, *85*, 89–99. http://dx.doi.org/10.1080/00220671.1991.10702818

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612–637. http://dx.doi.org/10.1037/0033-2909.128.4.612

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals* (1st ed.). Harlow, Essex: Longman Group. Retrieved from https://login.ezproxy.library.tufts.edu/login?url=https://search.proquest.com/docview/619961276?accountid=14434

Blumen, H. M., & Rajaram, S. (2009). Effects of repeated collaborative retrieval on individual memory vary as a function of recall versus recognition tasks. *Memory*, *17*, 840–846. http://dx.doi.org/10.1080/09658210903266931

Blumen, H. M., & Stern, Y. (2011). Short-term and long-term collaboration benefits on individual recall in younger and older adults. *Memory & Cognition*, *39*, 147–154. http://dx.doi.org/10.3758/s13421-010-0023-6

Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1118–1133. http://dx.doi.org/10.1037/a0019902

Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1563–1569. http://dx.doi.org/10.1037/a0017021

Congleton, A. R., & Rajaram, S. (2011). The influence of learning methods on collaboration: Prior repeated retrieval enhances retrieval organization, abolishes collaborative inhibition, and promotes post-collaborative memory. *Journal of Experimental Psychology: General*, *140*, 535–551. http://dx.doi.org/10.1037/a0024308

Foss, D. J., & Pirozzolo, J. W. (2017). Four semesters investigating frequency of testing, the testing effect, and transfer of training. *Journal of Educational Psychology*, *109*, 1067–1083. http://dx.doi.org/10.1037/edu0000197

Gatto, J. T. (2011, January 26). Does test-taking help students learn? [Letter to the editor]. The New York Times, p. A24.

Gaynor, J., & Millham, J. (1976). Student performance and evaluation under variant teaching and testing methods in a large college course. *Journal of Educational Psychology*, *68*, 312–317. http://dx.doi.org/10.1037/0022-0663.68.3.312

Gordon, L. T., & Thomas, A. K. (2014). Testing potentiates new learning in the misinformation paradigm. *Memory & Cognition*, *42*(2), 186–197. http://dx.doi.org/10.3758/s13421-013-0361-2

Healy, A. F., Jones, M., Lalchandani, L., & Tack, L. A. (2017). Timing of quizzes during learning: Effects on motivation and retention. *Journal of Experimental Psychology: Applied*, *23*, 128–137.

Iwaki, N., Nara, T., & Tanaka, S. (2017). Does delayed corrective feedback enhance acquisition of correct information? *Acta Psychol.*, *181*, 75–81. http://dx.doi.org/10.1016/j.actpsy.2017.10.005. PMID

Keys, N. (1934). The influence on learning and retention of weekly as opposed to monthly tests. *Journal of Educational Psychology*, *25*, 427–436. http://dx.doi.org/10.1037/h0074468

Kole, J. A., Healy, A. F., & Bourne, L. E., Jr. (2008). Cognitive complications moderate the speed-accuracy tradeoff in data entry: A cognitive antidote to inhibition. *Applied Cognitive Psychology*, *22*, 917–937.

Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology*, *29*, 210–212. http://dx.doi.org/10.1207/S15328023TOP2903_06

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., . . ., & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*, 51–57.

McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, *27*, 360–372. http://dx.doi.org/10.1002/acp.2914

McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition*, *1*, 18–26. http://dx.doi.org/10.1016/j.jarmac.2011.10.001

McDermott, K. B., Agarwal, P. K., D'Antonio, L., Roediger, H. L., III, & McDaniel, M. A. (2014). Both multiple-choice and short-answer quizzes enhance later exam performance in middle and high school classes. *Journal of Experimental Psychology: Applied*, *20*, 3–21. http://dx.doi.org/10.1037/xap0000004

Meade, M. L., Nokes, T. J., & Morrow, D. G. (2009). Expertise promotes facilitation on a collaborative memory task. *Memory*, *17*, 39–48. http://dx.doi.org/10.1080/09658210802524240

Mullet, H. G., Butler, A. C., Verdin, B., von Borries, R., & Marsh, E. J. (2014). Delaying feedback promotes transfer of knowledge despite student preferences to receive feedback immediately. *Journal of Applied Research in Memory and Cognition*, *3*, 222–229. http://dx.doi.org/10.1016/j.jarmac.2014.05.001

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005

Nguyen, K., & McDaniel, M. A. (2015). Using quizzing to assist student learning in the classroom: The good, the bad, and the ugly. *Teaching of Psychology*, *42*, 87–92. http://dx.doi.org/10.1177/0098628314562685

Pan, S. C., Cooke, J., Little, J., McDaniel, M. A., Foster, E. R., Connor, L. T., & Rickard, T. C. (2019). Online and clicker quizzing on jargon terms enhances definition-focused but not conceptually-focused biology exam performance. *CBE—Life Sciences Education*, *18*, 1–12.

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*, *144*, 710–756. http://dx.doi.org/10.1037/bul0000151

Pennebaker, J. W., Gosling, S. D., & Ferrell, J. D. (2013). Daily online testing in large classes: Boosting college performance while reducing achievement gaps. *PLoS ONE*, *8*, e79774. http://dx.doi.org/10.1371/journal.pone.0079774

Qualtrics (2018). Provo, UT.

Reisberg, D. (2013). *Cognition: Exploring the science of the mind: Sixth edition*. New York, NY: W.W. Norton & Company.

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. http://dx.doi.org/10.1037/a0026252

Sinha, N., & Arnold, L. (2015). Delayed, but not immediate feedback after multiple-choice questions increases performance on a subsequent short-answer, but not multiple-choice, exam: Evidence for the dual process theory of memory. *The Journal of General Psychology*, *142*, 118–134. http://dx.doi.org/10.1080/00221309.2015.1024600

Smith, M. A., & Karpicke, J. D. (2014). Retrieval practice with short-answer, multiple-choice, and hybrid tests. *Memory*, *22*, 784–802.

Szpunar, K. K., Jing, H. G., & Schacter, D. L. (2014). Overcoming overconfidence in learning from video-recorded lectures: Implications of interpolated testing for online education. *Journal of Applied Research in Memory and Cognition*, *3*, 161–164. http://dx.doi.org/10.1016/j.jarmac.2014.02.001

Wissman, K. T., & Rawson, K. A. (2016). How do students implement collaborative testing in real-world contexts? *Memory*, *24*, 223–239. http://dx.doi.org/10.1080/09658211.2014.999792

Wissman, K. T., & Rawson, K. A. (2018). Collaborative testing for key-term definitions under representative conditions: Efficiency costs and no learning benefits. *Memory and Cognition*, *46*, 148–157.

Wooldridge, C. L., Bugg, J. M., McDaniel, M. A., & Liu, Y. (2014). The testing effect with authentic educational materials: A cautionary note. *Journal of Applied Research in Memory and Cognition*, *3*, 214–221. http://dx.doi.org/10.1016/j.jarmac.2014.07.001

Received 16 October 2018;
received in revised form 20 December 2019;
accepted 21 December 2019
Available online xxx

Please cite this article in press as: Thomas, A. K., et al. Should You Use Frequent Quizzing in Your College Course? Giving up 20 Minutes of Lecture Time May Pay Off. *Journal of Applied Research in Memory and Cognition* (2020), https://doi.org/10.1016/j.jarmac.2019.12.005