Comparing Review Strategies in the Classroom: Self-Testing Yields More Favorable Student Outcomes Relative to Question Generation

E. Paige Lloyd1, Ryan J. Walker1, Molly A. Metz1, and Amanda B. Diekman1

Abstract
Although previous research has demonstrated that guided testing (i.e., self-testing) and question generation effectively increase retention compared to control methods, no work has simultaneously implemented both strategies in the classroom. In a semester-long study designed to maximize experimental control in a naturalistic setting, we adapted both review strategies for an introduction-level psychology course. We found that guided testing produced better student performance on exams than generating one’s own practice questions. Additionally, students evaluated guided testing more positively than question generation. These findings build upon previous guided testing and question generation work by showing that, in the context of an introductory classroom, guided testing is more effective and efficient than generating questions.

Keywords
study strategies, testing effect, generation effect, performance, student attitudes

Understanding which teaching methods facilitate student performance and generate positive student perceptions is of critical importance to instructors. Past research has demonstrated that the guided testing review strategy (often referred to as self-testing; e.g., Weinstein, McDermott, & Roediger, 2010), in which students take practice tests after learning material, facilitates learning (Andre & Anderson, 1978; Roediger & Karpicke, 2006a). Although this review strategy has been shown to be effective in educational and laboratory contexts (Carpenter, Pashler, & Vul, 2006; Leeming, 2002; McDaniel, Roediger, & McDermott, 2007), recent research has indicated that another review strategy, generating one’s own practice questions, is equally as effective in a laboratory setting (Weinstein et al., 2010). To our knowledge, no work has compared the effectiveness of these distinct review strategies in the classroom. Thus, the current study explored how guided testing and generating questions influenced performance in an introduction-level psychology class. Additionally, the current work examined student perceptions of these review strategies.

Testing Effect Versus Generation Effect
The positive effects of guided testing on retention are robust and well-replicated in the memory and educational psychology literatures (e.g., Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Bangert-Drowns, Kulik, & Kulik, 1991; Butler & Roediger, 2007; Karpicke & Roediger, 2007; Leeming, 2002; McDaniel, Roediger, et al., 2007; Nungester & Duchastel, 1982; Roediger & Karpicke, 2006a). Relative to rereading or reviewing material, taking short tests developed by the instructor or experimenter (without feedback) substantially increases retention (Weinstein et al., 2010), particularly long-term retention (Butler & Roediger, 2007; Nungester & Duchastel, 1982; Roediger & Karpicke, 2006a). The testing effect has widely been attributed to the fact that repeated retrieval from memory strengthens retrieval routes and results in increased elaboration (Bjork, 1975; Carpenter, 2009; McDaniel & Fisher, 1991; McDaniel & Masson, 1985; Roediger & Karpicke, 2006a, 2006b), similar to other generative study techniques (see Mayer et al., 2009, for a brief review). Notably, many of these memory experiments have been conducted outside of the classroom, making it difficult to generalize the findings to real-world classroom environments. The few guided testing studies that have been conducted in real or “simulated” classrooms have replicated the original effects.

1 Department of Psychology, Miami University, Oxford, OH, USA

Corresponding Author:
Ryan J. Walker, Department of Psychology, Miami University, 90 N. Patterson Ave., Oxford, OH 45056, USA.
Email: walkerrj@miamioh.edu
demonstrating that guided testing is a more effective review strategy than simply rereading (Batsell, Perry, Hanley, & Hostetter, 2017; Butler & Roediger, 2007; Einstein, Mullet, & Harrison, 2012; Kang, McDermott, & Roediger, 2007; McDaniel, Anderson, Derbish, & Morrisette, 2007; McDaniel, Roediger, et al., 2007; Roediger, Agarwal, McDaniel, & McDermott, 2011).

However, as most instructors realize, there are countless review strategies with claims to improved student performance and learning (e.g., Dunlosky & Rawson, 2015). Aside from asking students to participate in a practice test or quiz, which only requires students to quickly respond to questions, instructors can ask students to generate questions themselves (Bertsch, Pesta, Wiscott, & McDaniel, 2007; King, 1992). Asking students to generate their own questions requires additional processing not needed during guided testing (Weinstein et al., 2010). That is, while both types of review require cognitive effort, a hallmark of the generative theory of learning (Wittrock, 1989), generating questions requires students to search for, create, and synthesize material. Similar to the guided testing findings, studies have indicated that generating and synthesizing information results in greater retention compared to rereading or discussing the material (Foos, Mora, & Tkacz, 1994; Jacoby, 1978; King, 1989, 1991, 1992; McDaniel & Bugg, 2008; Slamecka & Graf, 1978).

Review techniques such as generating questions encourage students to process material in a deeper manner, which, by most accounts, should have a greater impact on learning compared to answering a series of questions (King, 1992; Robinson, 1970). In fact, the highest-level steps in the revised Bloom’s Taxonomy (Anderson, Krathwohl, & Bloom, 2001; Bloom, 1956), a six-step hierarchical model of learning objectives, are applying (i.e., using information in new situations; Level 3), analyzing (i.e., drawing connections; Level 4), evaluating (i.e., justifying a stance; Level 5), and creating (i.e., producing new work; Level 6). Due to the nature of asking students to generate their own questions, students are pushed to think critically, problem-solve, and reason, presumably reaching these higher-level learning objectives. In contrast, by asking students to simply respond to provided questions, instructors are not necessarily pushing students beyond the remember (i.e., recall facts; Level 1) and understand (i.e., explain ideas or concepts; Level 2) steps in Bloom’s Taxonomy. From this perspective, it seems plausible that asking students to generate questions in the classroom will lead to greater increases in performance and retention relative to guided testing.

In a recent series of studies, Weinstein and colleagues (2010) began to question this assumption. A series of laboratory studies compared the effectiveness of three review strategies: rereading, guided testing (i.e., self-testing), and generating test questions. As expected, generating questions and guided testing both led to significant increases in performance and retention compared to rereading the material. The authors also revealed that benefits of generation did not exceed or differ from those of guided testing (see also Denner & Rickards, 1987; Foos et al., 1994; Frase & Schwartz, 1975; Lehman & Lehman, 1984). Moreover, the authors demonstrated that generation, although equivalently effective, requires students to spend significantly more time on the review compared to guided testing. When selecting a review strategy, instructors may be interested in a strategy that offers learning benefits as well as efficiency for students. In this case, a guided testing review may provide greater utility than a generation review in a classroom setting.

Another potential boundary to the benefits of generation lies in the scaffolding necessary to employ this review effectively. Generation may prove effective for advanced students but fall flat for novices. Indeed, novice students lack the complex schemas developed by more experienced students (Kalyuga, Ayres, Chandler, & Sweller, 2003), and this lack of background knowledge is potentially problematic when considering the question generation review strategy. Specifically, novice students might not be able to elaborate on topics and generate questions in a productive way (King, 1992; Martin & Pressley, 1991), which might mean that students do not reach the higher critical thinking levels that make question generation effective in the first place. Instead, more directed instructional methods, such as guided testing, are more beneficial to these novice students (Kalyuga et al., 2003). The current study examined this very question: Is an in-class guided testing review more effective than an in-class question generating review in a course with novice students (i.e., an introductory-level psychology course)?

It is also important to note that many of the past studies on guided testing and question generation have primarily focused on retention effects or the cognitive mechanisms that drive these effects (e.g., Karpicke & Zaromb, 2010; McDaniel, Anderson, et al., 2007). Little work has explored student perceptions of these review strategies. How positively students view a pedagogy or study strategy (e.g., self-reported learning and level of engagement) is not only important for student motivation and instructor evaluations, but it also has implications for student performance (Armbuster, Patel, Johnson, & Weiss, 2009; Blasco-Arcas, Buil, Hernández-Ortega, & Sese, 2013; Carini, Kuh, & Klein, 2006; cf. Wesp & Miele, 2008). In Weinstein et al.’s (2010) work, participants perceived the question generation review to be more beneficial than the guided testing review (see also Agarwal et al., 2008; Leeming, 2002; Lyle & Crawford, 2011). The authors speculated that a likely reason for this finding was the substantial time difference to complete the tasks (i.e., question generation takes three times longer). In a real classroom, however, students are limited by time constraints. Therefore, in a classroom setting, an efficient task that allows students to quickly review material (i.e., guided testing) may result in more positive student perceptions relative to a more demanding task (i.e., question generation). However, it is also plausible that students will recognize that question generation requires more processing, which might lead students to perceive question generation more positively than guided testing. Thus, the second purpose of the current study is to further examine student perceptions of guided testing and question generation.
The Current Study

As outlined above, some evidence suggests that question generation will be more effective (relative to guided testing) for improving exam performance, whereas other evidence suggests the opposite. Thus, both hypotheses regarding which review strategy would be more effective were theoretically warranted, and we did not make a specific prediction. To control for the information students had available to them outside of class, each section had access to both the guided testing review (created by the instructors) and question generation review (created by the students) after completing the formal review during class. Data were not compiled or accessed until the completion of the course.

Materials and Design

Implementing these review strategies in the classroom can prove to be a challenge. As noted above, generation review strategies require much more time than testing review strategies (Weinstein et al., 2010). In controlled investigations of these review techniques, their differences necessitate that researchers choose between controlling for time spent reviewing and controlling for content or number of questions reviewed. In the current work, we were particularly interested in comparing review strategies that would be accessible to instructors wishing to implement them during class time.

To be confident that any effects would be due to the nature of the review and no other features of the course, instructor, or time in the semester, a 2 (Section: A vs. B) × 2 (Exam: 2 vs. 3) mixed design was employed, such that each section of students participated in each type of review, one type of review for each midterm. Exam review was not manipulated for the first exam (both sections employed the guided testing review), allowing us to control for initial student performance on the critical exam performance analysis.

Guided testing review. For the guided testing review, students responded to practice exam questions during the 55 min of class directly preceding an exam. Student responses were input via TurningPoint audience response system (Turning Technologies, Youngstown, OH). Students were presented with a question on the projector, and they were given approximately 45 s to respond with their personal clicker devices. Students were asked not to discuss the question or answer with fellow students. After all students responded to the question, the instructor presented a graph depicting the distribution of responses, and students were verbally given the correct answer. The instructors then provided the rationale as to why the answer was correct and why other answers were incorrect. Students were encouraged to ask questions if further explanation was desired. Following the review session, all practice questions from the review session were posted online (without answers) for use as an independent study resource. Each guided testing

Course Details

Advanced graduate students (one female, Section A; one male, Section B) taught the two sections of the course. Both instructors had one semester of experience teaching Introduction to Psychology. Prior to any experimental manipulations, we assessed student perceptions of their instructor’s warmth and competence as well as the students’ willingness to recommend their instructors to friends. We observed no differences between the two instructors (ps > .124). We repeated this procedure after data collection had completed and again observed no differences between the two instructors (ps > .133).

Students attended lecture for the course four days per week, 55 min per day. All lectures, activities, homework assignments, quizzes, and exams were standardized across sections. With the exception of names (e.g., reference to individual students or instructors) and class-specific details (e.g., meeting times), all materials were identical. Scripts for lectures were generated and discussed between instructors prior to every lecture. The Current Study

The Current Study

As outlined above, some evidence suggests that question generation will be more effective (relative to guided testing) for improving exam performance, whereas other evidence suggests the opposite. Thus, both hypotheses regarding which review strategy would be more effective were theoretically warranted, and we did not make a specific prediction. To control for the information students had available to them outside of class, each section had access to both the guided testing review (created by the instructors) and question generation review (created by the students) after completing the formal review during class. Data were not compiled or accessed until the completion of the course.

Materials and Design

Implementing these review strategies in the classroom can prove to be a challenge. As noted above, generation review strategies require much more time than testing review strategies (Weinstein et al., 2010). In controlled investigations of these review techniques, their differences necessitate that researchers choose between controlling for time spent reviewing and controlling for content or number of questions reviewed. In the current work, we were particularly interested in comparing review strategies that would be accessible to instructors wishing to implement them during class time.

To be confident that any effects would be due to the nature of the review and no other features of the course, instructor, or time in the semester, a 2 (Section: A vs. B) × 2 (Exam: 2 vs. 3) mixed design was employed, such that each section of students participated in each type of review, one type of review for each midterm. Exam review was not manipulated for the first exam (both sections employed the guided testing review), allowing us to control for initial student performance on the critical exam performance analysis.

Guided testing review. For the guided testing review, students responded to practice exam questions during the 55 min of class directly preceding an exam. Student responses were input via TurningPoint audience response system (Turning Technologies, Youngstown, OH). Students were presented with a question on the projector, and they were given approximately 45 s to respond with their personal clicker devices. Students were asked not to discuss the question or answer with fellow students. After all students responded to the question, the instructor presented a graph depicting the distribution of responses, and students were verbally given the correct answer. The instructors then provided the rationale as to why the answer was correct and why other answers were incorrect. Students were encouraged to ask questions if further explanation was desired. Following the review session, all practice questions from the review session were posted online (without answers) for use as an independent study resource. Each guided testing

Course Details

Advanced graduate students (one female, Section A; one male, Section B) taught the two sections of the course. Both instructors had one semester of experience teaching Introduction to Psychology. Prior to any experimental manipulations, we assessed student perceptions of their instructor’s warmth and competence as well as the students’ willingness to recommend their instructors to friends. We observed no differences between the two instructors (ps > .124). We repeated this procedure after data collection had completed and again observed no differences between the two instructors (ps > .133).

Students attended lecture for the course four days per week, 55 min per day. All lectures, activities, homework assignments, quizzes, and exams were standardized across sections. With the exception of names (e.g., reference to individual students or instructors) and class-specific details (e.g., meeting times), all materials were identical. Scripts for lectures were generated and discussed between instructors prior to every lecture. The Current Study

The Current Study

As outlined above, some evidence suggests that question generation will be more effective (relative to guided testing) for improving exam performance, whereas other evidence suggests the opposite. Thus, both hypotheses regarding which review strategy would be more effective were theoretically warranted, and we did not make a specific prediction. To control for the information students had available to them outside of class, each section had access to both the guided testing review (created by the instructors) and question generation review (created by the students) after completing the formal review during class. Data were not compiled or accessed until the completion of the course.

Materials and Design

Implementing these review strategies in the classroom can prove to be a challenge. As noted above, generation review strategies require much more time than testing review strategies (Weinstein et al., 2010). In controlled investigations of these review techniques, their differences necessitate that researchers choose between controlling for time spent reviewing and controlling for content or number of questions reviewed. In the current work, we were particularly interested in comparing review strategies that would be accessible to instructors wishing to implement them during class time.

To be confident that any effects would be due to the nature of the review and no other features of the course, instructor, or time in the semester, a 2 (Section: A vs. B) × 2 (Exam: 2 vs. 3) mixed design was employed, such that each section of students participated in each type of review, one type of review for each midterm. Exam review was not manipulated for the first exam (both sections employed the guided testing review), allowing us to control for initial student performance on the critical exam performance analysis.

Guided testing review. For the guided testing review, students responded to practice exam questions during the 55 min of class directly preceding an exam. Student responses were input via TurningPoint audience response system (Turning Technologies, Youngstown, OH). Students were presented with a question on the projector, and they were given approximately 45 s to respond with their personal clicker devices. Students were asked not to discuss the question or answer with fellow students. After all students responded to the question, the instructor presented a graph depicting the distribution of responses, and students were verbally given the correct answer. The instructors then provided the rationale as to why the answer was correct and why other answers were incorrect. Students were encouraged to ask questions if further explanation was desired. Following the review session, all practice questions from the review session were posted online (without answers) for use as an independent study resource. Each guided testing
review included approximately 30 questions (Exam 2 = 30; Exam 3 = 36).

**Self-generated review.** In the self-generation review (i.e., question generation), students spent the 55 min of class directly preceding an exam generating a collaborative study guide. Each student created three practice exam questions. To prevent students from generating only easy questions and to encourage deeper engagement with the material, students were urged to generate creative questions and were provided explicit instructions regarding the format and structure their questions should follow (see Appendix). Each student was assigned specific topics that would be covered on the test. The assignment of topics was employed to ensure a representative sample of material (mirroring that of the guided testing review) would be incorporated in the self-generation review guide. The instructor reviewed and approved each student’s questions once they finished the activity. Students were asked to revise a question if it was confusing, misleading, or did not fit the formatting requirements. Students were encouraged to ask questions if further explanation was desired. Duplicate questions and answers were not posted in the compiled study guide made accessible to students online.

**Exam performance.** Students completed three exams throughout the semester \(M_{\text{Exam1}} = 76.59, SD_{\text{Exam1}} = 13.75; M_{\text{Exam2}} = 81.05, SD_{\text{Exam2}} = 12.29; M_{\text{Exam3}} = 81.53, SD_{\text{Exam3}} = 12.74\), not including the final exam. Each exam was comprised of 50 multiple-choice or short answer questions worth a total of 100 points (6% of total grade). Exams were identical across sections.

**Student perceptions of review strategy.** We created a 5-item measure assessing student perceptions of the review strategies. For each item, students reported their agreement on a 9-point scale (1 = strongly disagree to 9 = strongly agree). Items included: “The instructor-guided review helped me on the exam”; “The instructor-guided review reduced my test-taking anxiety”; “The instructor-guided review improved my test-taking skills”; “The instructor-guided review provided a sense of control over my performance on the exam”; “The instructor-guided review increased my positivity toward collaboration with other students as a resource.” Items assessing the self-generated review specified “student-generated review” rather than “instructor-guided review.” Students completed this measure at two time points during the semester: following Exam 2 \(M_{\text{Exam2}} = 6.03, SD_{\text{Exam2}} = 1.53\) and following Exam 3 \(M_{\text{Exam3}} = 5.81, SD_{\text{Exam3}} = 1.69\). At both time points, Varimax-rotated principal components factor analyses indicated a single factor solution \(\alpha_{\text{Exam2}} = 3.45; \alpha_{\text{Exam3}} = 3.87\) and good reliability was observed \(\alpha_{\text{Exam2}} = .89; \alpha_{\text{Exam3}} = .93\).

**Procedure**

Preceding Exam 2, one section (Section B) employed the guided testing review and the other section (Section A) employed the self-generated review. Preceding Exam 3, the review strategies were switched across sections, such that Section A employed the guided testing review and the Section B employed the self-generated review. This allowed us to examine the effect of review strategy on performance independent of course section and instructor. Exam scores were linked with participant and section codes.

At the beginning of the course and before participating in a review session, all students provided informed consent. In the class meeting following each exam and prior to exam scores being released, students completed online anonymous surveys. Following Exam 1, students reported high school grade point averages (GPA) \(M = 3.67, SD = .34\), college GPA \(M = 3.25, SD = .51\), and the number of psychology classes they had previously taken \(M = .23; SD = .45\). Following Exams 2 and 3, students’ perceptions about the review strategies were assessed. Students also self-reported whether or not they attended the review during the previous class period. At the conclusion of every survey, students entered their participant and section codes.

**Results**

**Exam Performance**

To investigate whether review strategy impacted test performance, we conducted a 2 (class section: A vs. B) \(\times\) 2 (time: Exam 2 vs. Exam 3) mixed-model analysis of covariance (ANCOVA) on exam scores controlling for performance on Exam 1 (in which review strategy was not manipulated across class section). This analysis yielded only a section by time interaction, \(F(1, 82) = 5.03, p = .028, \eta^2_g = .06\). Although pairwise comparisons did not achieve significance \((ps > .057)\), the cross-over nature of this interaction is depicted in Figure 1. Both sections performed descriptively better on the exam with the guided testing review. That is, Section A performed better on Exam 3 (guided testing) than Exam 2 (self-generation), and Section B performed better on Exam 2 (guided
testing) than Exam 3 (self-generation). It is worth noting that even without Exam 1 as a covariate, this section by time interaction remains significant, $F(1, 84) = 5.25, p = .024, \eta^2_p = .06$. Moreover, this effect remains significant when controlling for high school GPA, college GPA, and previous psychology class experience, $F(1, 72) = 4.43, p = .039, \eta^2_p = .06$.

### Student Perceptions of Review Strategy

To investigate whether student perceptions of the review differed by review strategy, we conducted a 2 (class section: A vs. B) × 2 (time: Exam 2 vs. Exam 3) mixed-model analysis of variance (ANOVA) on student perceptions of the review strategies. This analysis yielded a marginal main effect of section, whereby Section A rated the review activities somewhat more favorably overall, $F(1, 84) = 3.60, p = .061, \eta^2_p = .04$. Critically, this effect was qualified by a section by time interaction, $F(1, 84) = 45.14, p < .001, \eta^2_p = .35$ (see Figure 2). Pairwise comparisons indicated that both sections evaluated the guided testing review more positively than the self-generation review. That is, Section A evaluated the review activity more positively on Exam 3 (guided testing) than Exam 2 (self-generation), $F(1, 84) = 14.98, p < .001, \eta^2_p = .15$, and Section B evaluated the review activity more positively on Exam 2 (guided testing) than Exam 3 (self-generation), $F(1, 84) = 31.29, p < .001, \eta^2_p = .27$. Thus, the guided testing review was evaluated by students in both sections as more helpful, as superior for reducing test-taking anxiety, as a greater contributor to test-taking skills, as providing a greater sense of control over performance, and as encouraging more positivity toward collaboration (relative to the self-generation review).

### Discussion

The results of the current study revealed a number of critical findings that contribute to our understanding of guided testing and question generation literatures. Most importantly, these data revealed that students in an introduction-level psychology course performed better following an in-class guided testing review relative to an in-class question generation review. Weinstein et al. (2010) demonstrated that guided testing and question generation are equally effective at facilitating retention in the lab. However, when the techniques are adapted for an actual classroom environment, in which review time is constrained and the students are expected to learn substantially more material, guided testing facilitates student performance to a greater extent. Second, these results demonstrate that students view the guided testing review more positively (e.g., more helpful, reduces anxiety, sense of control over performance) than the question generation review. This finding contrasts with Weinstein et al.’s (2010) work, which found that participants believed they would perform better following question generation than guided testing. Weinstein et al. (2010) provided participants with unlimited time in the lab to complete the study task, and participants took 3 times longer to complete the question generation review. The authors posited that the difference in time taken to complete the review might have driven this finding. However, in the current work, students had limited time to complete the review in the classroom. Guided testing is more efficient than question generation, meaning that students can review more material in the given study session. Thus, students may have rated the guided testing strategy more positively and as more effective (than question generation) because of the efficient nature of the strategy. Together, these findings establish that, in a classroom setting, guided testing is generally a more beneficial and efficient review strategy for both students and instructors.

Beyond contributing to review theory, the current work has notable implications for education broadly. First, instructors should find the results of this study to be applicable and generalizable to their classroom environments. Although experimental laboratory studies can demonstrate the effectiveness of a study strategy and establish a mechanism, classroom environments are complex. In classrooms, students have limited time and energy time to complete class activities, and they are expected to learn larger amounts of material over a semester (and that material is constantly being built upon and integrated). Moreover, students get anxious when taking real exams, and this anxiety can have a negative impact on performance (Cassady & Johnson, 2002). The control featured in any experimental study comes at the cost of some or all of these integral components of the student experience, making it difficult to generalize how students will respond to specific review techniques. It is therefore not surprising that educators have rarely integrated the experimental testing effect findings into classrooms (Kang et al., 2007; Matlin, 2002; McDaniel, Roediger, et al., 2007). Even when scholars attempt to generalize the experimental findings to the classroom, they do so by citing studies that were conducted in the lab or simulated classrooms (McDaniel, Roediger, et al., 2007). Thus, the current study provides novel insight into how students respond to one review technique relative to another over the course of a semester. Additionally, the current study highlights the efficiency of the guided testing review technique. When students are...
studying large quantities of information in a classroom, the efficiency of the guided testing strategy is only amplified. If instructors are looking for an effective review strategy that saves time and ensures students are retaining the material for an exam, then asking students to respond to comprehension questions is a better option than question generation.

Future Directions and Limitations

Some might wonder why guided testing ended up being a more effective review strategy than question generation, particularly when considering that question generation requires more in-depth processing, critical thinking, and synthesis of the material. The explanation to our findings may lie in an established phenomenon called the expertise reversal effect (Kalyuga, 2007; Kalyuga et al., 2003). Compared to expert students, students with less expertise might not possess the schemas necessary to synthesize information in a productive way during challenging tasks. When students do not possess these schemas, more complex tasks like question generation may force students to start using inefficient critical thinking strategies that are cognitively taxing. In contrast, expert students have developed the schemas that help them complete challenging tasks fluidly, without being cognitively overloaded. For this reason, novice students benefit more from the higher guidance methods, whereas expert students benefit more from the lower guidance methods (Kalyuga, 2007; Kalyuga et al., 2003). In the current study, we implemented these review strategies in classrooms of novice students, and a higher guidance review (i.e., guided testing) ended up facilitating performance to a greater extent than a less guided review (i.e., question generation). At more advanced levels, students are likely to have greater expertise, and instructors may have different learning goals for their students. In such courses, allowing time for self-generation of a wide range of questions may provide benefits. Had this study been conducted on advanced students, performance may have been facilitated more by question generation than guided testing, as the former strategy is less guided. Future work should examine how differing levels of prior knowledge influence the effectiveness of these review strategies.

Recent research by Karpicke and Zaromb (2010) also offers a memory-based explanation as to why guided testing facilitated performance to a greater extent than question generation. These authors identified an important distinction between the two study strategies. During guided testing, students are intentionally retrieving information that was stored during some class period or study session (called intentional retrieval). For each question, students are likely thinking back and attempting to reconstruct what they previously learned. In contrast, during question generation tasks, students do not necessarily need to think back and attempt to reconstruct what occurred in a previous class or study session. Although this might lead to in-depth processing of the material, at no point do students need to intentionally think back to a prior study session (Roediger & McDermott, 1993). Thus, question generation relies on what is called incidental retrieval conditions. This difference in retrieval mode (Tulving, 1983) may be responsible for the results of the current study (see Karpicke & Zaromb, 2010). Future work should continue to investigate how retrieval mode influences learning in the classroom.

A potential limitation to the current study is the amount of information each review allowed students to cover. As noted above, generating questions takes significantly more time than guided testing or similar review techniques (Andre & Anderson, 1978; Weinstein et al., 2010), which means students using that strategy cover less material. This difference becomes particularly apparent when both of these review strategies are implemented within a class period. In the current study, students using the guided testing review strategy covered approximately 30 questions, whereas students using the generation review strategy generated only three questions each. Although all students could access these questions outside of class, the difference in the amount of material covered during the review session may be responsible for the differences in student performance. We view this aspect of our design as both a limitation and strength. Given that we implemented these review strategies in the classroom, the amount of time students were given to work on a review activity was necessarily controlled. Had we asked students to complete these review strategies outside of class, which would have mitigated the above concern, the current study would have been less methodologically rigorous because it would not have been possible to ensure that students in each section actually completed the review activities. If instructors want to actively ensure students complete a review strategy, they may do so by implementing the review during a class period. When these reviews are implemented in class, the guided testing strategy is the more efficient strategy (relative to question generation) because it allows students to review substantially more material in a given session.

There are a few final limitations that are worth noting. First, given that all students participated in the guided testing review for Exam 1, students were more familiar with that review strategy. This increase in familiarity may have contributed to the positive effect that guided testing had on performance (relative to question generation). Second, how the instructors interacted with students differed between the review strategies. In the guided testing review, the instructor read the questions and then explicitly stated the answers. In the generation review, students generated their own questions and then received personalized feedback (for their questions and answer choices) from the instructor. Although students in both conditions received clarification from the instructor when needed, it is possible that differences in instructor interaction influenced the results. Third, the current study lacks a control condition, which is often represented by a rereading condition in the memory literature. Dozens of studies in the classroom and laboratory have demonstrated that the guided testing and question generation strategies improve performance and recall compared to rereading, discussing the material, or no review (Agarwal et al., 2008; Andre & Anderson, 1978; Bangert-Drowns et al., 1991; Batsell et al., 2017; Butler & Roediger, 2007; Carpenter et al., 2006;
and creative. Questions generated by all students will be compiled into a student generated “study guide.” However, the lack of a control condition in the current classroom study limits our ability to understand whether guided testing had a positive impact on performance, whether question generation had a negative impact on performance, or whether both forms of review had positive effects of different magnitudes. Future work might incorporate a control condition in the classroom to elucidate the direction of this effect. Incorporating a control condition in the classroom creates a potential ethical concern, however. As noted above, the memory literature often uses a rereading condition as the control. It is well established that rereading is not an effective study strategy relative to these other methods, such as guided testing (Dunlosky & Rawson, 2015). Thus, when the review strategy has implications for real student outcomes, such as a grade on an exam, asking students to use an ineffective review technique is ethically questionable.

**Conclusion**

Instructors may assume that study strategies requiring more critical thinking and material synthesis (e.g., question generation) afford greater student learning compared to more guided methods (e.g., guided testing). Although students do indeed benefit from these more challenging study strategies (King, 1992; McDaniel & Bugg, 2008), the current research indicates that guided testing offers greater performance and perception benefits for introduction-level students when incorporated into an in-class review. In sum, the current work builds upon the previous testing and generation effect research by demonstrating that guided testing is a more effective and efficient classroom review technique relative to question generation, at least in a classroom comprised of novice students.

**Appendix**

**Student Generated Review**

In this activity, you will generate three practice exam questions. Questions generated by all students will be compiled into a student generated “study guide.”

Try to make your exam questions clear, well thought out, and creative.

Follow the instructions below as you create your exam questions:

1. Questions must be multiple choice with four response options (A–D).
2. Make sure each question has only 1 correct answer.
3. Do NOT include the answers on the Google doc.

Once you have completed your questions:

1. Discuss them with your instructor.
2. Post them to the study guide Google doc, so that you and your classmates can use these practice exam questions to study for the exam.

**Question 1:** Generate an exam question based on the topic/concept you found to be the most interesting or the most difficult in this section of the course.

**Question 2:** Generate an exam question using people in your own life (or maybe even this class!) or based on an event from your own life.

**Question 3:** Generate a creative exam question on Module #______ (this number will change for each student in the class).

**Author’s Note**

Molly A. Metz’s in now affiliated to Department of Psychology, University of Toronto, Toronto, Ontario, Canada.

**Authors’ Contribution**

E. Paige Lloyd and Ryan J. Walker contributed equally to this research and are ordered alphabetically.

**Declaration of Conflicting Interests**

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The authors received no financial support for the research, authorship, and/or publication of this article.

**Notes**

1. These might be reasons why the testing effect literature has rarely been adopted by educators or integrated into classrooms (McDaniel, Roediger, et al., 2007).
2. Because the purpose of this study was to examine if one review strategy was more effective than the other, the course instructors were reluctant to manipulate the final exam review strategy. Manipulating review strategy on the final exam would have given one section a potentially unfair advantage over the other. In other words, assuming one review strategy resulted in better student performance than the other, one section of the course would have had the unfair advantage of using the more effective review strategy for two exams, whereas the other section would have only had the opportunity to use the more effective review strategy for one exam. Thus, both instructors implemented the self-generation and guided testing review strategies in hopes of offering students the greatest opportunity for success on the final exam.

**References**


Oxford University Press.


