

## **Classroom clickers offer more than repetition: Converging evidence for the testing effect and confirmatory feedback in clicker-assisted learning**

Amy M. Shapiro<sup>1</sup> and Leamarie T. Gordon<sup>2</sup>

*Abstract: The present study used a methodology that controlled subject and item effects in a live classroom to demonstrate the efficacy of classroom clicker use for factual knowledge acquisition, and to explore the cognition underlying clicker learning effects. Specifically, we sought to rule out repetition as the underlying reason for clicker learning effects by capitalizing on a common cognitive phenomenon, the spacing effect. Because the spacing effect is a robust phenomenon that occurs when repetition is used to enhance memory, we proposed that spacing lecture content and clicker questions would improve retention if repetition is the root of clicker-enhanced memory. In experiment 1 we found that the spacing effect did not occur with clicker use. That is, students performed equally on clicker-targeted exam questions regardless of whether the clicker questions were presented immediately after presentation of the information during lecture or after a delay of several days. Experiment 2 provided a more direct test of repetition, comparing test performance after clicker use with performance after a second presentation of the relevant material. Clicker questions promoted significantly higher performance on test questions than repetition of the targeted material. Thus, the present experiments failed to support repetition as the mechanism driving clicker effects. Further analyses support the testing effect and confirmatory feedback as the mechanisms through which clickers enhance student performance. The results indicate that clickers offer the possibility of real cognitive change in the classroom.*

*Keywords: clickers, feedback, clicker-assisted learning, knowledge acquisition*

Personal response systems, commonly called clickers, have become common in thousands of classrooms nationally. They allow instructors to assess comprehension and memory for material by posing a question to the class (usually multiple-choice) that students answer with remote devices they bring to class. Questions and answers take as little as a minute or two to present and collect, and voting results can be displayed instantly in a bar graph. Understandably, educators and researchers have been interested in the technology's educational effectiveness. Generally speaking, the majority of studies have shown that clickers are effective in boosting attendance and participation (Beekes, 2006; Poirier & Feldman, 2007; Shih, Rogers, Hart, Phillis, & Lavoie, 2008; Stowell & Nelson, 2007) and learning outcomes (Kennedy & Cutts, 2005; Mayer et al., 2009, Morling et al., 2008; Ribbens, 2007; Shapiro, 2009; Shapiro & Gordon, 2012). Few studies have explored the cognitive mechanism through which clickers increase retention of lecture content, however. The focus of the present work was to better

---

<sup>1</sup> University of Massachusetts Dartmouth, Psychology Department, 285 Old Westport Road, North Dartmouth, MA 02747-2300, [ashapiro@umassd.edu](mailto:ashapiro@umassd.edu)

<sup>2</sup> Tufts University

understand the cognition driving clicker effects. Specifically, the experiments presented here were designed to rule out repetition as the basis of clicker effects in fact-based learning, thereby supporting the hypothesis that the testing effect and feedback drive clicker effects.

There are several ways clickers may work to enhance memory for classroom material: (1) directing students' attention to material likely to be on exams, (2) repetition, and (3) the testing effect. The first possibility, that clickers "tip off" students about the instructor's judgment of important material, and therefore the content of exam questions, is a reasonable hypothesis. One might expect students to attend to those topics more in class and focus study effort on those topics. Greater attention in class and increased studying would both enhance exam performance. If the attention-grabbing hypothesis is correct, it would mean clicker questions do not directly enhance memory or learning. It would mean only that they are an effective means of directing learners' attention to particular topics. Repetition effects, the second possible way clickers enhance memory for classroom material, make the justification for clicker use similarly debatable. Repetition can be accomplished through online resources, readings or other assignments outside of class, without using any class time and at no cost to students. If clicker effects are attributable instead to the last possibility, the testing effect, it would indicate they have unique benefit in the classroom. Writing clicker questions and integrating them with class lectures does require a modest time investment. Once that is completed, however, clicker questions require little class time to administer, correct and enter to grade sheets. Indeed, the entire sequence of presentation, response, grading, recording and feedback all happens within seconds. As such, if clickers are due to the testing effect rather than repetition or attention-grabbing it would mean they offer unique benefit of enhanced learning during class time with very little investment of time or money.

Shapiro and Gordon (2012) were able to rule out attention-grabbing and found modest support for the testing effect during clicker use in a live classroom. In their study, a series of exam questions were targeted over the semester in two classes. Half the items in one class were targeted with clicker questions when the information was taught in class. The other half of the questions was targeted by attention alerts. They assigned the same items to the opposite conditions in the other class. This counterbalanced the assignment of each question to the experimental and control conditions, and created a situation in which each item served in both the clicker and attention conditions. Students did not get clicker questions about the information assigned to the attention condition. Instead, they were told that the information was very important and would be covered on the next test. The relevant information on the PowerPoint slide was also highlighted in red and was animated to flash. At the end of the semester students were given a survey that asked what directed their decisions about what to study. In spite of the fact that they reported studying the information targeted by the alerts more than that targeted by clicker questions, students performed as well or better on questions when a clicker question was offered. In short, even when attention was explicitly drawn to specific information in class and studied more outside of class, answering a clicker question had an equal or greater effect on exam performance. That study did not rule out attention-grabbing as a contributing factor to clicker effects, but it did provide strong evidence that it is unlikely to be the sole source of clicker effects. The authors argued that the testing effect also underlies clicker effects.

Shapiro and Gordon (2012) were not able to rule out the possibility of repetition effects as the mechanism underlying clicker effects, however. Because they compared a clicker group to a no-clicker control group that was exposed to one presentation of the material, repetition is confounded with clicker use. Indeed, the majority of studies that report clicker effects compare

clicker use with no clicker use, with no control for repetition effects (e.g., Mayer et al. 2009; Morling et al, 2008; Shapiro, 2009). At present, then, it is unclear whether the testing effect or simple repetition effects are driving clicker effects in the classroom. The present study was designed to address this question. We sought to determine whether the learning outcomes observed with clicker use are attributable to repetition. Before explaining the methodology, we provide a brief review on the research that explains these phenomena.

### **The Testing Effect and Repetition Learning**

Karpicke, Roediger and others have documented that testing memory can enhance later recall or recognition better than an equivalent amount of additional study (Butler, Karpicke, & Roediger, 2007; Carrier & Pashler, 1992; Karpicke & Roediger, 2007a, 2007b, 2008; Roediger & Karpicke, 2006a; Szpunar, McDermott, & Roediger, 2008). In what has become the classic paradigm for investigating the testing effect, Thompson, Wenger, and Bartlings (1978) gave one group 3 study sessions followed by a delayed test (SSST). Another group studied the same information once and was then tested 3 times (STTT), the final test serving as the dependent measure after a 48 hour delay. On the final test, the SSST group forgot 56% of the material, as opposed to just 13% by the STTT group. This basic effect has been demonstrated using free recall (Jacoby, 1978; Szpunar et al., 2008), short-answer (Agarwal, Karpicke, Kang, Roediger, & McDermott, 2006) and multiple-choice (Duchastel, 1981; Nungester & Duchastel, 1982) tests and has been demonstrated with memory for word lists (Karpicke & Roediger, 2007a; Tulving, 1967), paired-associates (Allen, Mahler, & Estes (1969) and text (Nungester & Duchastel, 1982; Roediger and Karpicke, 2006a).

The cognition underlying the testing effect is not fully understood but some hypotheses have emerged and are currently under investigation. One possibility is that repeated testing creates conditions in which information is over-learned, a position argued by Thompson et al. (1978). Over-learning is an unlikely explanation of clicker effects, as it is improbable that offering a single clicker question in class can lead to over-learning. A more likely possibility is that testing strengthens the pathways leading to a stored memory more than additional study does (Bjork, 1975). Since study can be very passive (e.g., re-reading text passages or lecture notes), the more active nature of generating responses or comparing multiple-choice alternatives could reasonably offer greater opportunity for such enhancement. In other words, individuals are engaging in an activity that requires greater concentration during testing than some forms of study. Indeed, Bjork and Bjork (1992) have argued that there is a positive relationship between the level of effort required during testing and the strength of memory. As such, the effect may be a form of depth of processing ( Craik & Lockhart, 1972).

Alternatively, testing may generate new routes to the memory trace, thus multiplying possible access points to the material (McDaniel & Masson, 1985). When memories are formed, information about the context and activities relevant to the material are also formed. Testing offers new perspectives and links to the information that may be sensitive to different memory cues than the connections formed during study. The latter possibility would take advantage of encoding specificity, as a pathway generated through testing is likely to be more easily accessed during later testing. An excellent and more extensive review of the testing effect is provided by Roediger & Karpicke (2006b).

Although the mechanisms underlying the testing effect are not fully understood, numerous investigations have demonstrated that the effect seems to be enhanced by feedback

(e.g., Butler & Roediger, 2007; Hattie & Timperley, 2007; Kulhavy, 1977; Pashler, Cepeda, Wixted, & Rohrer, 2005; Sassenrath & Gaverick, 1965; Thorndike, 1913). Feedback can be confirmatory or corrective, and there is evidence that both types enhance later test performance (Butler, Karpicke, & Roediger, 2007; Kluger & DeNisi, 1996; McDaniel et al., 2007; Vojdanoska et al., 2010). Because clickers allow instructors to provide feedback with a simple button click within seconds of voting, feedback is widely used among clicker-adopting instructors. As a consequence, feedback is an important facet of clicker use to consider when questioning the reasons underlying clicker-mediated learning effects, particularly the testing effect.

It is important to note that the testing effect has been demonstrated in many experiments that did not employ feedback (see Kang, McDermott, & Roediger, 2007, experiment 1; Marsh, Agawal, & Roediger, 2009; Roediger & Karpicke, 2006a), so while there is the potential for the contribution of feedback effects during clicker-based learning, some other mechanism unique to testing appears to be working with or in addition to feedback. A study by Kang et al. (2007) underscores this point. After reading journal articles, subjects took either short answer or multiple-choice tests prior to a final memory test. Subjects did better on the final test when they took preliminary multiple-choice tests. When feedback was offered on the preliminary tests (in experiment 2), however, students taking the short answer tests did better on the final test. In sum, testing improved learning in Kang et al.'s study, but the addition of feedback altered something about the mechanism involved. The results are highly suggestive of some sort of interaction between the memory processes relevant during testing and feedback.

In spite of the fact that testing, especially with feedback, has been shown to enhance performance on tests more than study repetition, mere re-exposure to material alone can improve learning. The more times a student is exposed to a bit of information, the greater the likelihood he or she will retain it (e.g., Ebbinghaus, 1913; Raney, 2003; Scarborough, Cortese, & Scarborough, 1977; Tulving, 1967). As such, it is certainly possible that clicker questions may improve retention for classroom content merely by re-exposing students to the material. In other words, clicker effects may simply be repetition effects, and that is a potential criticism of any experiment that demonstrates clicker effects by comparing clicker use with a no-clicker control. Thus, it is important to rule out repetition as the cause of clicker effects in order to strengthen the argument for classroom clickers as effective and worthwhile pedagogical tools.

## The Present Study

Shapiro and Gordon (2012) concluded that the testing effect, not attention-grabbing, is responsible for enhanced learning with clickers in their experiment. Because they compared clicker groups to non-clicker control groups, as do most published studies on the topic, clicker use was confounded with repetition in their investigation. In the present two-experiment study we tested whether clicker effects are due, at least in part, to repetition.

Experiment 1 takes advantage repetition learning in order to determine the role of repetition in clicker effects. Specifically, if repetition is a significant source of clicker effects, clicker use should be subject to the *spacing effect*. The spacing effect (also called *distributed learning*) refers to the phenomenon in which rehearsal or re-exposure to material results in greater memory when a period of time is allowed to intervene between presentations (Benjamin & Tullis, 2010; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Glenberg, 1979; Hintzman, 1974). If clicker questions are more effective when offered after a delay of several days, it will

indicate the questions are likely serving as a method of repeating exposure to class material. If the spacing effect is not evident, it will indicate that repetition is unlikely to be a significant factor in clicker effects.

In experiment 2, we compared a clicker group that received a single presentation of the material and a subsequent clicker question to a group that received a second presentation of the material in place of the clicker question. Because Shapiro and Gordon (2012) have found evidence against attention-grabbing as the reason for clicker effects, failure to support repetition in the present study would provide converging evidence that clicker effects are most likely attributable to the testing effect. We also took advantage of the clicker data to perform a secondary analysis on clicker performance to learn something about the role of feedback in clicker effects.

## **Experiment 1**

The experiment was designed to determine whether the clicker learning effects demonstrated in prior studies are subject to the spacing effect, and thus attributable to repetition effects. We designed experiment 1 to compare exam question performance when clicker questions were asked immediately after in-class presentation of the material and when clicker questions were asked after a delay. If the spacing effect is in evidence, subjects should score higher on test items when clicker questions were offered 2-5 days after the material was taught in class, as compared with the same clicker questions offered the same day. Finding a spacing effect would indicate that clicker effects may be attributed, at least in part, to repetition. If the spacing effect does not emerge in the data, it would indicate that either feedback or the testing effect leads to cognitive change that can't be attributed to simple rehearsal. For this reason, an analysis of clicker question performance was conducted to determine the role of feedback apart from repetition.

## **Method**

### **Subjects**

Four hundred students enrolled in two sections of general psychology at the University of Massachusetts participated in the study. Students participated as part of their normal coursework, and earned participation points by correctly answering in-class questions. They ranged from freshmen to seniors and represented a range of disciplines offered at the institution. IRB approval was sought prior to beginning the study and a waiver was granted.

### **Materials and Procedure**

The class covered 11 topics in general psychology and was taught as a typical lecture course with demonstrations and multimedia integrated into many of the lectures. PowerPoint presentations were projected onto a movie theater-sized screen. In-class clicker questions were integrated into the presentations, with individual slides dedicated to single questions. The iClicker system was used to allow students to make their responses to clicker questions.

Students were required to purchase their clickers (for \$20-40, depending on whether they were new or bundled with the required text book).

Sixteen test clicker question/test item pairs were used as stimuli in the present study. Each clicker question was written to tap the same information as its targeted exam question. All clicker and exam questions were multiple-choice and were taken from Shapiro and Gordon (2012). The clicker question/test item pairs were spread throughout the semester, and across the four exams administered during the semester. Performance on the exam questions was the dependent variable.

All the targeted exam questions were included in the exams for both classes. The clicker question written for each targeted exam question was also given to each class. The timing of the clicker question presentation was manipulated as the within-subjects independent variable. When assigned to the “immediate” condition, clicker questions were given in class directly after the material was presented and any student questions were answered. When assigned to the “delayed” condition, the questions were given at the start of another class meeting, 2-5 days after the material was taught. Half the items were included in each condition for one class, with the other half included in the opposite condition for the other class. As such, each of the 16 experimental items was included in both the immediate and delayed conditions, and each subject contributed data to both conditions. Presentation of the relevant course material was the same in both conditions; the information was included on a PowerPoint slide. Identical “filler” clicker questions targeting material unrelated to the experimental items were offered to both classes, with the experimental items mixed randomly among them. Between 1-5 clicker questions (filler and experimental) were asked in class each day. The instructor projected the clicker questions onto the screen after soliciting and answering any questions from the students. Students were given 30-90 seconds to answer each question and a bar chart showing the percentage of the class to respond with each option was projected to provide feedback after voting was closed.

**Exam and clicker question validation.** Because a simple, no-clicker control condition would not allow discrimination between clicker and repetition effects, which is the purpose of this investigation, a no-clicker group was not included. For that reason, it was important to establish that the materials used in the present study do induce a basic learning effect. As mentioned, the sixteen clicker questions, and the corresponding exam questions for which they were written, were taken from Shapiro and Gordon (2012). The clicker question written for each exam question probed the same basic information as the test question, but was still unique. In their study, Shapiro and Gordon implemented a counterbalancing strategy wherein each of two classes was given clicker questions for half the targeted exam questions. For the other half of the questions, subjects were given no clicker question. For half of those in the control condition (see experiment 1), no special treatment was given to the material in class. For the other half, however, students were told the material was important and would be on the test (see experiment 2), creating a very conservative test of clicker learning effects. The methodology controlled for both item and subject effects, as each exam question was used in the control and clicker conditions and each subject contributed data to both conditions. Half the stimuli in the present experiment were taken from Shapiro and Gordon’s experiment 1 and half from experiment 2. Thus, in order to establish that the item subset chosen for the present study does produce the basic clicker learning effect, the analysis from that experiment was re-run including only the subset of items chosen for the present study. Analyzed by subjects, a paired t-test

revealed a significant effect of clickers on performance,  $t(234) = 5.62$ ,  $p < .0001$ ,  $d=.37$ . Students scored a mean of 68.9% (SD=18.7) correct on items when no clicker question was offered and 76.8% (SD=18.1) correct when a question was offered, more than an 11% performance increase. The results were also significant when analyzed by items,  $t(15) = 4.29$ ,  $p < .001$ ,  $d=1.08$ , with items answered correctly by 69.4% (SD=12.1) of subjects when placed in the control condition and 76.0% (SD=10.8) answering the same questions correctly when clicker questions were asked, an increase of almost 10%. Again, this is a very conservative test of the stimuli because half the items in the control condition were identified to students as material that would be on the test. In spite of the warning, clicker questions still significantly boosted exam performance.

Other measures of the stimuli were taken to ensure stimulus validity. Two independent content experts provided validation ratings of the stimuli. Both are professors of psychology that routinely teach introductory psychology. They rated each clicker and exam question on a 7-point scale for the following dimensions: (1) overall quality of the question, (2) relevance of the information targeted by the clicker/exam item pairs to the content and goals of an introductory psychology course, (3) the relationship between each clicker item and each exam question. The questions used in the experiment all scored a minimum rating and minimum mean of 5.0 by each rater on questions 1 and 2. The clicker/exam pairs met the same criteria on survey question 3. The relationship ratings between clicker questions and exam questions which were *not* intended as pairs were also analyzed. It was important that unpaired items were actually unrelated to ensure clicker questions were not enhancing memory for exam questions for which they were not written. All unrelated clicker/exam question pairs used in the present experiment scored a maximum rating of 2.0 among reviewers and had a mean rating of 1.5. The low ratings established the unlikelihood of “spillover” effects. That is, clicker questions were unlikely to affect performance on exam questions for which they were not intended.

## Results and Discussion

Students who withdrew early from the course, those with attendance lower than 60%, and those who missed more than one exam were excluded from the data analysis. These students provided insufficient data for the within-subjects comparisons or were insufficiently exposed to the independent variable. The deletions yielded a total of 283 subjects in the analysis. Moreover, individual exam question data were removed from the analysis for students who were absent from class the day the targeted content was presented. Missing those critical classes meant missing the targeted content as well as their immediate clicker questions. Also, effects of the delayed clicker questions would be difficult to interpret for those cases. A maximum of 16 exam questions per subject was possible and these deletions resulted in a mean of 13.1 per subject. Out of a maximum of 283 student scores for each question, the deletions resulted in a mean of 229.6.

Paired *t*-tests were performed to compare performance between the immediate and delayed conditions. The results did not reveal evidence of a spacing effect. When analyzed by subjects, there was no significant difference between performance on exam items when targeted by immediate ( $M = 67.5$ ,  $SD = 24.1$ ) or delayed ( $M = 70.0$ ,  $SD = 21.8$ ) clicker questions,  $t(282) = 1.73$ ,  $p > .05$ . No significant difference between the immediate ( $M = 67.4$ ,  $SD=9.3$ ) and delayed ( $M = 69.8$ ,  $SD=11.7$ ) conditions was revealed in the item analysis,  $t(15) = 1.04$ ,  $p > .05$ . The mean discrimination index for the exam questions was 50.4.

Since there was no spacing effect, the data argue against repetition as a significant mechanism underlying clicker effects. If repetition isn't driving the effect, what is? A clue to the relevant processes may be gleaned by examining clicker question performance in the immediate versus delayed conditions. It makes intuitive sense that students would perform better on immediate clicker questions, as the information needed to answer the questions correctly has just been presented in lecture. In light of the fact that students performed equally well on later exam questions regardless of clicker question timing, however, if students did perform better on the immediate versus delayed clicker questions it would suggest corrective feedback is being used to improve test performance to some extent. Paired *t*-tests by items, comparing clicker question performance between immediate and delayed conditions revealed just that. Students scored a significantly higher percent correct on immediate clicker questions ( $M = 94.7$ ,  $SD=9.1$ ) than delayed ( $M = 83.2$ ,  $SD=15.8$ ),  $t(281) = 10.99$ ,  $p < .0001$ ,  $d = .65$ . The same result was found when analyzed by items, with the same clicker questions answered correctly more often when asked in the immediate condition ( $M=94.7$ ,  $SD=5.1$ ) than in the delayed condition ( $M=82.0$ ,  $SD=18.3$ ),  $t(15) = 2.85$ ,  $p < .01$ ,  $d = .72$ . Not only are the *t*-tests significant, but the effect sizes are quite robust. Despite such clear differences between immediate and delayed clicker performance, exam performance was not affected by conditions. As such, it stands to reason students were able to make some use of their performance feedback in the delayed condition to improve test performance.

The clicker performance analysis provides only indirect evidence about the effect of corrective feedback, however. A more direct test is possible by comparing exam question performance when the clicker questions were answered correctly versus incorrectly. If feedback is a primary factor in clicker effects, students should score equally on exam questions regardless of clicker performance as long as they are given feedback, as they were in the present study. If there is a significant difference, it would mean the effect of corrective feedback is limited and unlikely to account for the entire effect. To run this test, all subjects and questions in the delayed clicker condition were combined to create groups based on clicker performance. Because clicker performance was quite high in the immediate condition (95%), there were insufficient incorrect responses to compare with the correct responses, so the analysis was done only on the delayed clicker questions. Moreover, since exam question performance was deleted when the critical content lecture was missed, there are no cases in the immediate clicker condition in which students attended the content lecture but missed the clicker questions. The delayed condition, however, provides an important comparison group. That is, students who attended the critical content lecture but were not exposed to the delayed clicker question.

The limitations of corrective feedback effects are seen when performance is compared on test items for which students correctly versus incorrectly answered the corresponding clicker questions, or did not see the clicker questions. The mean of the 1422 exam questions included in the analysis, for which the corresponding clicker questions were correctly answered<sup>3</sup>, was 72%. For the 286 exam questions, for which the corresponding clicker questions were incorrectly answered, the mean score was 63% correct. There were 191 unanswered, delayed clicker questions across subjects that did attend the critical content lecture (in other words, students who received the content in class but did not see the clicker question) and the mean score on the corresponding exam questions was 59%. Although the effect size was quite small, the difference was significant,  $F(2, 1896) = 9.83$ ,  $p < .0001$ ,  $\eta^2 = .01$ . A Scheffe's posthoc analysis revealed that

---

<sup>3</sup> With 16 items and 283 subjects, there were 2284 possible clicker responses in the immediate and in the delayed conditions. The number in the analysis is lower due to student absences.

exam question performance was significantly higher in the case of correctly answered clicker questions than incorrectly answered clicker questions,  $p < .05$  (two tailed) and in the case of correct versus missed clicker questions,  $p < .05$  (two tailed). The difference between exam question performance based on incorrect versus missed clicker questions was not significant,  $p > .05$  (two tailed).

If corrective feedback were a primary mechanism through which clicker effects worked, there should be little or no significant difference on exam performance based on clicker performance. More importantly, incorrectly answered clicker questions should yield better performance than getting no clicker question at all. After all, if students are using clicker questions primarily to gain corrective feedback on their performance, one would expect to see evidence of widespread self-correction on the exam questions. The significant performance advantage by students getting the answer correct, in addition to the comparable exam performance of students getting a clicker question wrong and those unexposed to it, suggests corrective feedback was not particularly useful for students getting clicker questions wrong. The large differences in sample sizes and the rather low effect size, however, warrant caution about the strength of this conclusion.

## **Experiment 2**

The purpose of experiment 2 was to provide converging evidence with experiment 1 that repetition is not the major source of clicker learning effects. The advantage of the methodology used in experiment 1 was that the presentation of immediate and delayed clicker questions seemed natural to students within the context of a live classroom. Taking advantage of the spacing effect in this way, however, only provided indirect evidence of the role of repetition. Experiment 2 addressed the question more directly by comparing exam question performance after the presentation of clicker questions or information repetition. Moreover, since the main evidence refuting repetition effects in experiment 1 was a nonsignificant result, experiment 2 was also designed to provide positive evidence (i.e., a significant statistical result) in support of our hypothesis.

## **Method**

### **Subjects**

Three hundred twenty students enrolled in two sections of General Psychology at the University of Massachusetts participated in the study. Students participated as part of their normal coursework, and earned participation points by correctly answering in-class questions. They ranged from freshmen to seniors and represented all five colleges across campus. IRB approval was sought prior to beginning the study and a waiver was granted.

### **Materials and Procedure**

The same materials and procedure the same procedure was used as in experiment 1, but with one change. Instead of half the exam questions being targeted with delayed clicker questions in each semester, half were targeted with a second, immediate presentation of the material. In the clicker and repetition conditions, the same slide was used to present the information for the first

time. In the clicker condition a clicker question followed the slide. In the repetition condition a second PowerPoint slide that presented the relevant information in a slightly different way from the first was presented in lieu of a clicker question. In this way, the effect of a second, novel presentation on exam question performance could be compared with the effect of a clicker question. A sample stimulus set from each condition is provided in Appendix A. In both conditions, the targeted information was presented verbally along with an accompanying slide. (In the Appendix A example, the targeted information was the role of the hypothalamus in hormone regulation.) In the repetition condition, the information was repeated with a new visual aid, while in the clicker condition students answered a question in lieu of seeing the second slide.

## Results and Discussion

Students that withdrew early from the course, those with attendance lower than 60%, and those that missed more than one exam were excluded from the data analysis. This yielded a total of 290 students in the analysis. Paired *t*-tests were performed to compare performance between the clicker and repetition conditions. The results indicated significantly better performance in the clicker condition ( $M = 61.2$ ,  $SD = 21.6$ ) than in the repetition condition ( $M = 56.2$ ,  $SD = 20.6$ ) when analyzed by subjects,  $t(289) = 3.417$ ,  $p = .001$ ,  $d = .20$ . The effect was also significant when analyzed by items,  $t(15) = 2.419$ ,  $p = .029$ ,  $d = .60$ , with students performing better on items when the relevant content was presented with a clicker question ( $M = 60.7$ ,  $SD = 10.4$ ) rather than with a second presentation ( $M = 55.2$ ,  $SD = 12.0$ ).

The results of experiment 2 converge with those of experiment 1 to support the hypothesis that clicker questions do not enhance retention of classroom material merely because they act as a second presentation of information. The 5-point increase in subject performance (from 52.2 to 61.2) in the subject analysis represents a performance increase of 8.9%. The effect size is rather small, however. The 5.5-point increase in the item analysis represents a 10% increase and a moderate effect size, however. While these results can't rule out any role of repetition in clicker effects, they do provide compelling evidence that repetition is not the major source of the effect.

## General Discussion and Conclusions

Shapiro and Gordon (2012) reported evidence that clicker effects are not attributable to drawing students' attention to certain material. That study was not able to rule out repetition effects as an underlying cause of clicker-enhanced learning, however. The present study addressed that possibility and demonstrated that repetition is unlikely to be a major contributor to the effect. In doing so, it provides converging evidence with Shapiro and Gordon that the testing effect is likely to underlie clicker-enhanced learning.

In a secondary analysis of experiment 1, we tried to determine whether feedback has a role in clicker effects, since feedback is an important variable in the testing effect. The conclusions we were able to draw from those analyses are suggestive of some role of feedback, but do not paint a clear picture. The delayed clicker group performed worse on clicker questions than the immediate group but performed equivalently on exam questions, suggesting that corrective feedback helped. However, a comparison of exam question performance when students correctly versus incorrectly answered the clicker questions revealed students performed

better on exam questions when they got clicker questions right. Indeed, students answering the clicker question incorrectly performed only as well on the exam questions as students that were not exposed to the clicker question at all. These results suggest corrective feedback had a weak effect on exam performance. Any conclusions drawn from the latter result, however, are mitigated by the rather low effect size. On balance, then, the present results are suggestive of some role of corrective feedback in clicker-based learning. That conclusion is compatible with the large literature on the role of feedback in the testing effect. Certainly, feedback should be an important area for future inquiry.

Regardless of the feedback question, the results do converge with Shapiro and Gordon (2012) to support the conclusion that the testing effect is the most likely mechanism underlying clicker effects. The notion of testing itself causing cognitive change is supported by the extensive work of Karpicke and colleagues (e.g., Karpicke & Roediger, 2007a; 2008) on the testing effect. As Bjork (1975) suggests, the act of retrieving memories may strengthen the memory trace. Moreover, it may create new routes to memories that are more easily invoked during exams, with the context common to testing situations acting as a retrieval cue.

The present experiment was designed to test clicker use for enhancing fact-based learning alone. As such, the results do not support clicker use for problem-solving, application, or deep-level understanding of the material. Within the context of fact-based learning, however, the present results are of practical importance for educators and students. As such, we can offer some concrete suggestions for effective use of clickers in the classroom. Specifically, we suggest that important factual content be targeted with clicker questions. The questions should be written specifically to require memory retrieval of the targeted information. We also suggest the questions be worded clearly and in a way that maximizes students ability to correctly answer the questions. After all, if the testing effect is at the heart of clicker-enhanced learning, the goal should be to encourage students to correctly recall the correct information from memory, thereby activating the testing effect.

Finally, clickers seem to invoke cognitive change in the classroom that is unique. If clicker effects were attributable to repetition or attention-grabbing, their value might be dubious. After all, there are many avenues through which to provide repetition or enhance attention inside and outside the classroom. Having demonstrated that clicker use affects cognitive change attributable to the testing effect (and quite possibly to feedback, as well) the present results support clickers as a unique and valuable pedagogical classroom tool. Given the relatively low cost in terms of classroom time and equipment expense, the evidence in support of their educational benefit suggests they do offer real value to students and instructors.

## References

- Agarwal, P. K., Karpicke, J. D., Kang, S. K., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22, 861-876. doi:10.1002/acp.1391
- Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning & Verbal Behavior*, 8(4), 463-470. doi:10.1016/S0022-5371(69)80090-3

Beekes, W. (2006). The "Millionaire" method for encouraging participation. *Active Learning in Higher Education: The Journal of the Institute for Learning and Teaching*, 7, 25-36. doi:10.1177/1469787406061143

Benjamin, A., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive Psychology*, 61, 228-247. doi:10.1016/j.cogpsych.2010.05.004

Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R.L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium* (pp. 123-144). Hillsdale, NJ: Erlbaum.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, & R. Shiffrin (Eds.), *From Learning Processes to Cognitive Processes: Essays in Honor of William K. Estes Volume 2* (pp. 35-67). Hillsdale, NJ: Erlbaum.

Butler, A. E., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, 13, 273-281.

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19, 514-527. doi:10.1037/1076-898X.13.4.273

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20, 633-642. doi:10.3758/BF03202713

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J.T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132, 354-380. doi:10.1037/0033-2909.132.3.354

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, 11, 671-684. doi:10.1016/S0022-5371(72)80001-X

Duchastel, P. C. (1981). Retention of prose following testing with different types of tests. *Contemporary Educational Psychology*, 6, 217-226. doi:10.1016/0361-476X(81)90002-3

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. (H. A. Ruger & C. E. Bussenius, Trans.). New York: Teachers College Press.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, 7, 95-112.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112. DOI: 10.3102/003465430298487

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola symposium* (pp. 77–97). Potomac, MD: Erlbaum.

Jacoby, L. L. (1978). On interpreting the effects of repetitions: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*, 649-667. doi: 10.1016/S0022-5371(78)90393-6

Kang, S. H. K., McDermott, K. B., & Roediger, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, *19*, 528-558. doi: 10.1080/09541440601056620

Karpicke, J. D., & Roediger, H. L. (2007a). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151-162. doi: 10.1016/j.jml.2006.09.004

Karpicke, J. D., & Roediger, H. L. (2007b). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 704-719. doi: 10.1037/0278-7393.33.4.704

Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*, 966-968. doi: 10.1126/science.1152408

Kluger, A. & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*, 254-284 doi: 10.1037/0033-2909.119.2.254.

Kennedy, G. E., & Cutts, Q. I. (2005). The association between students' use of an electronic voting system and their learning outcomes. *Journal of Computer Assisted Learning*, *21*, 260-268. doi: 10.1111/j.1365-2729.2005.00133.x

Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, *47*, 211-232. doi:10.2307/1170128

Marsh, E. J., Agarwal, P. K., & Roediger, H. L. (2009). Memorial consequences of answering SAT II questions. *Journal of Experimental Psychology: Applied*, *15*, 1-11. doi:10.1037/a0014721

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Bulger, M., Campbell, J., Knight, A., & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, *34*, 51–57. doi:10.1016/j.cedpsych.2008.04.002

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 371-385. doi:10.1037//0278-7393.11.2.371

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology, 19*, 494-513. doi:10.1080/09541440701326154

Morling, B., McAuliffe, M., Cohen, L., & DiLorenzo, T. (2008). Efficacy of personal response systems (“clickers”) in Large, Introductory Psychology Classes. *Teaching of Psychology, 35*, 45-50. doi:10.1080/00986280701818516

Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology, 74*, 18-22. doi:10.1037/0022-0663.74.1.18

Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 3-8. doi:10.1037/0278-7393.31.1.3

Poirier, C. R., & Feldman, R.S. (2007). Promoting active learning using individual response technology in large introductory psychology classes. *Teaching of Psychology, 34*, 194-196. doi:10.1080/00986280701498665

Raney, G. (2003). A context-dependent representation model for explaining text repetition effects. *Psychonomic Bulletin & Review, 10*, 15-28. doi:10.3758/BF03196466

Ribbens, E. (2007). Why I like clicker personal response systems. *Journal of College Science Teaching, 37*, 60-62.

Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255. doi: 10.1111/j.1467-9280.2006.01693.x

Roediger, H. L., & Karpicke, J. D. (2006b). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210. doi:10.1111/j.1745-6916.2006.00012.x

Sassenrath, J.M., & Garverick, C.M. (1965). Effects of differential feedback from examinations on retention and transfer. *Journal of Educational Psychology, 56*, 259-263. doi:10.1037/h0022474

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception & Performance, 3*, 1-17. doi:10.1037//0096-1523.3.1.1

Shapiro, A. M. (2009). An empirical study of personal response technology for improving attendance and learning in a large class. *Journal of the Scholarship of Teaching and Learning, 9*, 13-26.

Shapiro, A.M., & Gordon, L.T. (2012). A controlled study of clicker-assisted memory enhancement in college classrooms. *Applied Cognitive Psychology, 26*, 635–643. doi: 10.1002/acp.2843.

Shih, M., Rogers, R., Hart, D., Phillis, R., & Lavoie, N. (2008, April). *Community of practice: The use of personal response system technology in large lectures*. Paper presented at the University of Massachusetts Conference on Information Technology, Boxborough, MA.

Stowell, J., & Nelson, J. (2007). Benefits of electronic audience response systems on student participation, learning, and emotion. *Teaching of Psychology, 34*, 253-258. doi:10.1080/00986280701700391

Szpunar, K. K., McDermott, K. B., & Roediger, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1392-1399. doi:10.1037/a0013082

Thompson, C. P., Wenger, S. K., & Bartlings, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 210-221. doi:10.1037/0278-7393.4.3.210

Thorndike, E. L. (1913). *Educational psychology: Vol. 1. The original nature of man*. New York: Columbia University.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall verbal learning. *Journal of Verbal Learning and Verbal Behavior, 6*, 175-184. doi:10.1016/S0022-5371(67)80092-6

Vojdanoska, M.; Cranney, J., & Newell, B. (2010). The testing effect: The role of feedback and collaboration in a tertiary classroom setting. *Applied Cognitive Psychology, 24*, 1183-1195. DOI: 10.1002/acp.1630

## Appendix

### Appendix A. Sample Stimulus Set.

Sample item in the clicker and repetition conditions, reproduced in grayscale.

TARGETED EXAM QUESTION:

Which brain structure exerts considerable influence over the secretion of hormones throughout the body?

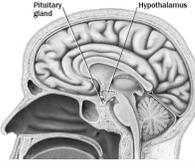
- A. the hypothalamus
- B. the amygdala
- C. the hippocampus
- D. the thalamus

### EXPERIMENTAL CONDITION Clicker

### Repetition

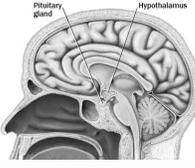
### FIRST PRESENTATION

**Hypothalamus**



• Located deep in the brain  
• Controls hormones and regulates a number of functions

**Hypothalamus**



• Located deep in the brain  
• Controls hormones and regulates a number of functions

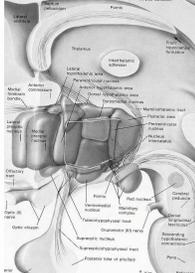
### SECOND PRESENTATION

iClicker Question

Which of the following is NOT a function of the hypothalamus?

1. Hormone regulation
2. Thirst
3. Sleep
4. All of these are hypothalamus functions

**Hypothalamus**



- Temperature regulation
- Controls hormones (endocrine system)
- Sexual activity
- Hunger
- Thirst
- Sleep