Spring 2014

Experimental and Theoretical Analysis of Cognitive Processes Underlying Clicker Use in STEM Education

Erica Rozbruch
University of Colorado Boulder

Experimental and Theoretical Analysis of Cognitive Processes Underlying Clicker Use in STEM Education

Erica V. Rozbruch

March 21, 2014

Thesis Advisor: Alice F. Healy (Department of Psychology & Neuroscience)

Thesis Committee: Ryan Bachtell (Department of Psychology & Neuroscience), Sanyu Mojola (Department of Sociology), and Matt Jones (Department of Psychology & Neuroscience)

University of Colorado at Boulder

Department of Psychology & Neuroscience

Acknowledgements

I would like to express my gratitude to Dr. Alice Healy, my research advisor, for her continuous support from the beginning of the project, the weekly supervision meetings, computing statistical analyses, and the critiques and suggestions she gave me on my honor's thesis. Without her guidance, supervision, and love for research, this honor's thesis would not have been possible.

I would like to offer my thanks to Matt Jones for developing the experimental designs with Alice Healy and for computing and interpreting some of the statistical data analyses.

I wish to thank Lindsay Anderson and Dan Corral for creating the programs used in both of the experiments, for training me on how to utilize the different computer programs, for constructing the experimental materials, for data tabulation, and for testing some of the participants.

Lastly, I would like to thank all of the research assistants in the Jones Lab who tested a majority of the participants for Experiment 2.

Experimental and Theoretical Analysis of Cognitive Processes Underlying Clicker Use in STEM Education

The clicker technique has been gaining popularity in higher education, but it is still unknown how to use the clicker technique most effectively. The present experiments investigate this innovative approach to the teaching and learning of science, technology, engineering, and mathematics (STEM) disciplines.

The clicker technique offers an interactive method of teaching that allows teachers to gauge how well the students understand the presented material so the teacher can alter the lecture, accordingly. The clicker technique is comprised of instructors during lecture giving multiple-choice questions about lecture material, in which students can respond via a hand-held device, the clicker. After all the students have submitted their answers, via a clicker, both the instructors and students receive immediate feedback in the form of a frequency distribution per each multiple-choice answer.

Several studies have established links between clicker use and learning outcomes in college classrooms. Mayer et al. (2008) found that a class section that received clicker questions earned higher grades in the course than sections of the same class that did not receive clicker questions. Donovan (2008) paired clicker questions with exam questions and found there was improvement on the exam questions that were originally presented as clicker questions. Smith et al. (2009) found that a combination of discussion and use of the clicker enhanced students' understanding of lecture material. Although there are positive impacts that the clicker technique has on learning, the underlying cognitive mechanisms are still unknown. Caldwell (2007) reviewed research regarding the use of clickers in classrooms and concluded that there is not enough evidence to make any

scientific conclusions about what causes these benefits of learning in the clicker technique. The present experiments were conducted in efforts to optimize teaching and learning in higher education classrooms, as well as to reveal the underlying cognitive mechanisms that are present in the clicker technique.

Effects of Delay (Timing)

Much research has investigated the effects of temporal spacing of learning, finding that spaced practice leads to better learning than massed practice (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Research has shown that students have a difficult time paying attention to a lecture for longer than 15 minutes, and breaking up the lecture by asking clicker questions can be an effective way to increase attention, and therefore, retention of material (Middendorf & Kalish, 1996). These previous studies have focused on the retention of material that is taught over a long period of time, but we hypothesize that spacing also has benefits for teaching in that it can improve the reliability of assessment. Specifically, it is hypothesized that delayed clicker questions will be better predictors than immediate clicker questions of eventual exam performance, because the delayed questions will better match the exams in terms of the cognitive processes they depend on. This prediction can also be viewed from the perspective of depth of processing (Craik & Lockhart, 1972) because delayed testing involves deeper processing and requires more elaborative rehearsal than does immediate testing (Craik & Watkins, 1973).

Prior research has been conducted for recall in immediate and delayed conditions with regards to a list-before-last paradigm; participants recall a list of words that was not most recently presented, but the one before it. Participants in the delayed condition who

had to isolate the list in the testing condition learned how to cope with a switch in context from the studying to the testing phase (Jang & Huber, 2008). Jang and Huber looked at context retrieval and context similarity; participants in the "recall" condition had to recall information in a different context, whereas participants in the "no test" condition had to recall information in a similar context (Jang & Huber, 2008). The ability to carry learned information from the study phase to the testing phase is essential, as participants needed to do this in order to be successful at the recall task. However, Jang and Huber (2008) found that only recall, not recognition, effectively drives context change.

In the first experiment of the present study, using a laboratory version of the clicker technique (Anderson, Healy, Kole, & Bourne, 2011, 2013), time was varied relative to study material at which clicker questions were asked. Three between-subjects conditions were compared: Immediate 1, Immediate 2, and Delayed. In each condition, there were eight blocks during learning, with each block comprising a different subtopic for study. In the Immediate conditions, clicker questions pertained to material from the current block. In contrast, in the Delayed condition, clicker questions pertained to material from one block earlier (as in the list-before-last paradigm). In the Delayed condition, participants were shown two blocks of information before being tested on the first block of information. As there were eight blocks of information, participants were only tested on seven blocks of information, due to the delay. There were two different Immediate conditions (Immediate 1, Immediate 2) that controlled for temporal positions of studying versus testing**.** In both Immediate 1 and 2, participants were shown a block of information and then tested immediately after each block of information was presented. In Immediate 1, participants were tested on the first 7 blocks of information, but not the

last block, to keep consistency with the Delayed condition in terms of the temporal spacing of the *study* phase. In Immediate 2, participants were not tested on the first block of information, yet they were tested on the last seven blocks of information to keep consistency with the Delayed condition in terms of the temporal spacing of the *testing* phase (see Figure 1). All participants were trained and tested on the same amount of information.

Effects of Response Type (Recall vs. Recognition)

Clicker questions are given to students in an "A-E" multiple-choice format and require students to recognize the correct answer out of five alternative answers. An experiment that implemented a comparable clicker system suggested that utilizing a multiple-choice format is useful for fact memorization, but not necessarily for more complex questions (Guthrie & Carlin, 2004). It is well known, however, that efforts to recall can facilitate later retention relative to recognition (e.g., Hogan & Kintsch, 1971). The second experiment of the present study compares between subjects the standard *recognition-only* clicker procedure with a *recall-only* procedure and a *recall-recognition* procedure developed by Smith and Healy (1998). In this procedure, clicker questions are presented alone, and students are asked to covertly generate answers before the choices are presented (see Kang et al., 2011, for a similar procedure). The students are then given the correct answer as feedback after both generating their own answer and recognizing the correct answer out of alternative answers. Kang et al. (2011) found that wrong guesses do not hurt students' acquisition of correct information as long as feedback is provided shortly after**.** Furthermore, we adopted the procedure of forced guessing in the present experiment because it is found that requiring someone to guess does not impair

subsequent fact learning when feedback is provided shortly after, even when participants selected incorrect answers (Kang et al., 2011). In all three conditions, recognition performance (on a multiple-choice test) was examined at both the posttest and retention test. Expectations based on Kintsch's (1970) two-stage model of memory (in which recall includes a generation stage followed by an editing stage; see also Anderson & Bower, 1972, for a similar generate-recognize theory) are that posttest performance will be better when clicker questions require recall, because recall requires a process of retrieval that strengthens the memory representation of retrieved items.

Because recall, but not recognition, requires memory retrieval, it is also hypothesized that clicker questions based on recall will be better predictors of both posttest and retention test performance. However, this hypothesis goes against the transfer appropriate processing principle (Morris, Branford, & Franks, 1977), by which the multiple-choice posttest questions more closely resemble those given in the studying/teaching round of the recognition-only condition than those given in the studying/teaching round of the recall-recognition condition and recall-only condition (see, e.g., Balota & Neely, 1980**).**

Experiment 1

Using a laboratory model of the clicker technique, Experiment 1 explored the effects of the clicker in three conditions: two involving quizzes immediately after studying and one involving delayed quizzes. Testing occurred after a brief distractor task and two days later. Questions on the tests were presented either in a general format or in the specific format seen at study.

Method

Participants. Fifty-six undergraduate students taking a course in General Psychology at the University of Colorado at Boulder participated in this experiment. Of the 56 participants, we used data from only 48 participants. Four of the remaining students completed Part II of the experiment more than two days later. Experimenter errors led to three participants being tested in the wrong condition. One student never came back to finish Part II. Of the 48 participants, 28 were women and 20 were men.

Materials. In this two-part experiment, every participant was trained and tested on an Apple iMac desktop computer with the procedure programed in Python. In the middle of Part I, we handed out a distractor task, which consisted of a packet requesting participants to read two passages and detect target letters or letter sequences in the passages.

Procedure. Each student was tested individually in separate rooms. Participants were first told to read the instructions shown on the computer screen (see Appendix A). These instructions informed them that they would be viewing several sets of eight facts about different types of plants, and that they would be tested on their ability to retain facts about those plants. During Part I, each participant was trained in one of three conditions (Immediate 1, Immediate 2, or Delayed) and in one of four counterbalancing conditions. The current study utilized a fact-learning task, which consists of 64 facts about eight different plant categories (trees, herbs, etc.), following Anderson et al. (2013). Each plant category had eight different exemplars whose names were fictitious. All of the fictitious plant names were orthographically regular nonwords generated from actual plant names in a given plant category. The facts were presented as sentences, each including a plant category, a description, and a name. Each fact for a given fictitious plant

is true for a given, matched real plant. Each fact for each plant exemplar existed in two forms: a general form (e.g., "A tree that comes from Asia is the *Pawthra*") and a specific form (e.g., "A tree that is native to southern India is the *Pawthra*"*)*. The italicized word (the fictitious plant name) is what participants were tested on during the four-alternative multiple-choice quiz. The facts in training were always in the specific form. The correct answer was provided during feedback for each multiple-choice quiz question. After the training phase, participants completed the distractor task. After the distractor task, they took one of two versions of a posttest, with half of the questions in each block in general format and half in specific format on each posttest version. Across the two posttest versions, a given fact occurred in both formats. Half of the participants saw one version of the posttest, and the other half saw the alternate version of the posttest, depending on their counterbalancing condition. Counterbalancing condition also determined the placement of the specific correct answer ("A", "B", "C", or "D") for a given question on the clicker test, posttest, and retention test, with the constraint that each letter was the correct answer twice in each block of eight questions for each test version. The posttest served the role of a midterm exam and consisted of questions based on the plant facts that they had just learned. Each participant came back 2 days later to be tested for Part II. This retention test represented the final exam. If participants had originally seen a question in the general form on the posttest, they saw the question in the specific form on the retention test, and vice versa.

Results

In an analysis of accuracy, there was a significant interaction of experimental condition and phase, $F(4,90) = 5.03$, $MSE = 0.042$, $p = .001$, reflecting worse

performance in the Delayed condition than in the Immediate conditions on the clicker test. However, there was comparable performance on the posttest and retention test (see Figure 2).

There was also a significant main effect of phase, $F(2,90) = 123.05$, $MSE = 0.042$, $p < .001$, indicating declining performance from training (clicker test) ($M = .682$, $SE = .012$) to the posttest ($M = .553$, $SE = .014$) to the retention test ($M = .433$, $SE = .012$). Performance also depended on the specific quiz (i.e., the category type), $F(6, 270) = 6.74$, $MSE = 0.048$, $p < .001$. The means and standard errors of the seven quizzes were as follows: Quiz 1 ($M = .525$, $SE = .019$), Quiz 2 ($M = .477$, $SE = .020$), Quiz 3 ($M = .555$, $SE = .022$), Quiz 4 ($M = .594$, $SE = .022$), Quiz 5 ($M = .561$, $SE = .021$), Quiz 6 ($M = .625$, $SE = .021$), Quiz 7 ($M = .553$, $SE = .023$).

An analysis limited to the posttest and retention test revealed that participants performed worse on the retention test ($M = .432$, $SE = .013$) than on the posttest ($M = .559$, $SE = .016$), $F(1,45) = 60.76$, $MSE = 0.078$, $p < .001$, signifying that they forgot information after the two-day delay. There was also a significant effect of the format of facts asked during the posttest and retention test. Overall, participants were less accurate when they were tested on general facts, ($M = .419$, $SE = .013$) than on specific facts ($M = .571$, $SE = .015$), $F(1,45) = 114.00$, $MSE = 0.029$, $p < .001$. Because the participants were all presented with facts given in a specific format during the training phase, participants who were tested on facts given in a general format during testing were required to transfer their knowledge. The discrepancy between performance for the general and specific questions is reduced from the posttest [general ($M = .451$, $SE = .019$), specific ($M = .666$, $SE = .021$)] to the retention test [general ($M = .387$, $SE = .018$),

specific ($M = .477$, $SE = .018$)]; the interaction of test time and format was significant, $F(1,45) = 7.68$, $MSE = 0.073$, $p = .008$.

A marginally significant interaction between the format of questions (specific and general) and condition (Immediate 1, Immediate 2, and Delayed), $F(2,45) = 3.09$, $MSE = 0.029$, $p = .055$, is due to relatively high performance by the Immediate 1 condition on the general questions and relatively low performance by the Delayed condition on the specific questions [Immediate 1: general ($M = .462$, $SE = .026$), specific ($M = .592$, $SE = .026$); Immediate 2: general ($M = .379$, $SE = .021$), specific ($M = .581$, $SE = .025$); Delayed: general ($M = .417$, $SE = .023$), specific ($M = .540$, $SE = .026$). Thus, the Delayed condition showed less cost of transfer (from specific to general questions) than did either of the Immediate conditions.

In an analysis of the correlations between clicker test performance and performance on the posttest, the estimated increase in probability correct at test attributable to being correct on the corresponding clicker question is 19.9% for Delayed, 10.8% for Immediate 1, and 9.8% for Immediate 2. A contrast comparing the Delayed condition to the other two yields $F(1,2528) = 5.20$, $p = .023$. For the retention test, the estimated increase in probability correct at test attributable to being correct on the corresponding clicker question is 15.6% for Delayed, 8.4% for Immediate 1, and 14.9% for Immediate 2 (see Figure 3). The contrast between conditions is not statistically significant in this case.

In summary, clicker performance (by individual subjects on individual questions) predicts posttest performance better in the Delayed condition than it does in the Immediate conditions.

Discussion

Experiment 1 was conducted to evaluate the best way to use clickers in a classroom to gauge the students' understanding of the material and determine when teachers should ask questions during a lecture. Participants were tested in a lab setting by giving them fictitious facts about trees, herbs, wild flowers, etc. They were tested after a distraction task and two days later, with these tests representing a midterm exam, and the latter, a final exam. Although participants did worse in the Delayed condition than in the Immediate condition during training, we can use the Delayed condition test as a better predictor of eventual exam performance. These implications are extremely valuable in a classroom setting where students and teachers utilize clickers. From our data, it is suggested that teachers test their students on blocks of information after additional material is presented, representing the Delayed condition. Teachers should not ask clicker questions directly after they teach a block of information, because although students may perform better initially, this is not a good determinant of how well they will perform on the midterm or final exam. It is extremely important to maximize the efficiency and productivity associated with the use of clickers in a classroom setting. Further experiments will test other hypotheses in order to determine the best way to utilize this way of facilitation and teaching. The clicker technique provides an ongoing assessment of class knowledge. In this experiment, the knowledge of students was assessed by gauging their understanding of the material and predicting their subsequent performance to maximize classroom efficiency. Delayed assessment is more predictive of long-term retention, because of cognitive processes underlying immediate and delayed retrieval.

There is a stronger correlation between clicker and exam performance for material in the Delayed condition than material in the Immediate conditions.

Experiment 2

Using a laboratory model of the clicker technique, Experiment 2 explored the effects of the clicker response type: *recognition-only, recall-only*, or *recall-recognition*. All participants were tested under the recognition condition, but training condition was varied by fixed rotation. Testing occurred both immediately after a study phase and a week later. A pilot study was conducted to test an initial version of a subset of stimuli utilized in Experiment 2 (see Appendix B).

Method

Participants. One hundred and forty-five undergraduate students taking a course in General Psychology at the University of Colorado at Boulder participated in this experiment during Fall 2013 in order to receive credit.

Materials. In this two-part experiment, every participant was trained and tested on an Apple Mac mini computer with the procedure programmed in MATLAB. The 12 slides used during the study phase were taken directly from the first lecture of an introductory course titled "Psychological Statistics and Research Methods" at the University of Colorado at Boulder. These teaching slides were primarily relevant to the basics of research methods. During Session 1, each participant completed a study phase that included three clicker quizzes, depending on condition, including 16 questions in total. Participants were then given a posttest, and one week later, a retention test during Session 2. The posttest included 24 questions in total; half of the clicker questions (8), half of the test questions (8), and half of the Transfer 1 questions (8) were presented to a

given subject, whereas another subject saw the other half of the questions. The clicker questions were directly associated with the teaching slides that the participants were supposed to learn during the study phase; a definition was presented and the participant selected the term that fit the definition. The test questions were also based directly off of the teaching slides during the study phase, but this time participants were given a term and had to correctly identify the definition that went with the term. The Transfer 1 questions consisted of various scenarios, and subjects were prompted to pick the term(s) that best fit the scenario. The retention test, which occurred one week later, included the other half of the clicker, test, and Transfer 1 questions that the participants did not receive during the study phase. Additionally, every participant was tested on two other lists of transfer questions, each including 14 questions. The participants then took the retention test consisting of 52 multiple-choice questions. It should be noted that every participant was tested on the same set of questions. The Transfer 2 questions consisted of various scenarios where the participants had to determine whether or not a confounding variable was present during the experiment. The Transfer 3 questions also consisted of various scenarios where participants had to recognize which confounding variable was present and had to explain how to change the study to remove the confounding variable.

Twenty-nine participants who did not partake in the original study took the retention test measure as a way to gauge a baseline level of pre-existing knowledge without learning and testing during the study phase.

Procedure. Twenty-nine students came into the laboratory and took the retention test measure on a computer with all the questions without prior training or testing. This test counted as a baseline measure for the control group. Participants in the experimental

group were first instructed to read the instructions shown on the initial computer screen. These instructions informed the subjects that they would be participating in a memory task. In this between-subjects design, participants were in one of the three training conditions: *recall-only, recognition-only*, or *recall-recognition*. The study phase consisted of three different sections, and in each section, instructions were first given on a screen informing the participants how many teaching slides they had to study in the specified amount of time. All participants were allowed to study the teaching slides for the same amount of time. In the clicker quiz of the recall-only condition, participants were asked a question and prompted to use the keyboard to type in their answer on the screen. They could then proceed to the next question by hitting "enter," but they were required to generate some response before continuing. In the recognition-only condition, participants were to complete a clicker quiz in the format of a multiple-choice test. Participants were given a question with five answer options (i.e., "a," "b," "c," "d," and "e"), and they were prompted to type in the letter corresponding to the correct answer and to hit "enter" to continue. Similar to the recall-only condition, these students were asked to type in some response before proceeding. In the recall-recognition condition, the clicker quiz questions were presented alone, and students were asked to type in their answer choice accordingly before the choices were presented. Then, they answered the question in multiple-choice format. In both instances, they were asked to type in some response before proceeding.

In Section 1 of the study phase, following two preliminary slides, participants had to study two teaching slides for at least 5 minutes. They could go back and forth between these two slides for 5 minutes, and afterwards, the participants were allowed to proceed

to take the clicker quiz, or they could take an extra 2 minutes to study the material further. The first clicker quiz consisted of three questions. Sections 2 and 3 consisted of four slides each with a minimum of 10 minutes and a maximum of 14 minutes to study the slides. The clicker quiz after Section 2 consisted of eight questions, whereas the quiz after Section 3 consisted of five questions. The clicker quizzes immediately followed the teaching slides in all three sections, and the correct answer was given immediately following each of the participants' responses in all three conditions (following their answer to the multiple-choice question in the recall-recognition condition).

After the study phase was complete, the participants took a posttest, which consisted of 24 multiple-choice questions. The questions were selected from a total of 48 questions, 16 of which came directly from the study phase (clicker questions), 16 of which were new posttest questions (test questions), and the other 16 of which came from a third list of posttest questions (Transfer 1 questions). Questions were counterbalanced in each of the three training conditions. Feedback was not given during either of the posttests in all three conditions. In counterbalancing Condition 1, participants saw the even-numbered questions of the clicker test, but saw the odd-numbered questions for both the test questions and Transfer 1 questions during the posttest. For the retention test, in counterbalancing Condition 1, participants saw the odd-numbered questions of the clicker test, but saw the even-numbered questions for both the test questions and Transfer 1 questions. This selection was reversed for counterbalancing Condition 2. In the retention test subjects also saw all 28 of the two remaining types of questions (14 questions each from Transfer 2 and Transfer 3). The order of all questions was

randomized for both the posttest and retention test with a different random order shown to each subject on each test.

Results

An initial analysis was conducted on the test data (posttest and retention test) for the clicker, test, and Transfer 1 questions. On average, participants across all conditions performed better on the posttest (M = .762, SE = .010) than on the retention test (M = .692, SE = .010), due to forgetting across the one-week delay; there was a main effect of test time, $F(1, 142) = 34.68$, $MSE = 0.032$, $p < .001$, and comparable performance between the memory conditions of recall-only (M = .713, SE = .012), recognition-only (M = .736, SE = .013), and recall-recognition (M = .732, SE = .012), $F(2,142) < 1$. There was a significant main effect of question type as participants across all conditions performed best on clicker questions (M = .824, SE = .010), while performing worse on test (M = .701, SE = .012) and Transfer 1 questions (M =. 656, SE = .013), $F(2, 284) = 114.90$, $MSE = 0.019$, $p <.001$.

A separate analysis, including the experimental and control condition, was conducted on Transfer questions 2 and 3 (the far transfer questions) which were included in the retention test. Performance was better on the Transfer 2 questions (M =. 317, SE = .014) than on the Transfer 3 questions (M = .256, SE = .011); the main effect of far transfer question type was statistically significant, $F(1, 172) = 17.62$, $MSE = 0.014$, $p <.001$.

This analysis also compared the experimental and control participants. Participants performed better in the experimental condition (M =.298, SE =.010) than in the control condition (M =.229, SE =.023). The difference between the experimental and

control conditions is statistically significant with $F(1,172) = 5.55$, $MSE = 0.041$, $p = .0197$, demonstrating that participants learned from the material about experimental design and were able to apply their knowledge to successfully answer far transfer questions above the chance (guessing) level (.2).

When comparing the control test with the experimental retention test on all five question types, there was a significant effect on performance [control ($M = .388$, $SE = .019$), experimental ($M = .534$, $SE = .010$); $F(1,72) = 29.07$, $MSE = 0.088$, $p < .001$]. This effect illustrates that participants in the experimental group retain information over time and are able to apply their knowledge and answer questions that require them to elicit transfer mechanisms, even those requiring far transfer. Students learn specific facts and are able to transfer and generalize these facts.

Discussion

Experiment 2 was conducted to assess the most effective way to utilize clickers in a classroom setting by determining how students should respond via the clicker. The clicker technique works by presenting a question to the students and offering five different answer choices in an "A-E" multiple-choice format. However, clicker questions might be presented in a different format during test time. By training participants in three different experimental conditions (recall-only, recognition-only, and recognition-recall), we were able to determine which method will produce the most desirable outcomes during eventual exam performance. Overall, students performed better during the posttest, suggesting that over time, students forgot information. This information could be vital to teachers as they implement the clicker technique in their classroom. Teachers should assign students supplemental materials throughout the course of the class to

encourage practice and repetition of the material. Teachers should always require students to guess even if they are not sure of the correct answer because incorrect guessing will not harm their later performance.

Although the three different experimental conditions did not differ significantly, all of the participants were able to learn from the clicker training and transfer their knowledge to more difficult questions. The transfer appropriate processing principle suggests that participants should perform better on eventual exam performance if they were originally trained under the recognition-only phase, because they were tested under that same test format. However, it was hypothesized that participants would perform better in the recall conditions because they relied on their ability to formulate an answer, which requires more effort than recognizing a correct response. Because there is not a significant effect of training condition, it is hypothesized that these two contradictory hypothesized effects cancelled each other out.

Overall, participants performed better on Transfer 2 questions than on Transfer 3 questions. Both of these sets of questions tested participants' ability to transfer their knowledge, but as Transfer 2 questions only required the participants to recognize a confounding variable, Transfer 3 questions required the participants to explain their reasoning by choosing the appropriate way to remove the confounding variable from options "A-E". Although the participants' performance was low on Transfer 2 and 3 questions, they still performed better than the control group, suggesting that the training phase led the participants to learn information to be able to transfer their knowledge to more difficult questions.

Future analyses will examine the predictability of clicker performance on posttest and retention test performance for each training condition to determine which condition provides the best predictors of eventual exam performance, as in Experiment 1 (see Figure 3).

General Discussion

The clicker technique is a useful tool that can be easily implemented in a classroom and facilitates students to learn material, recognize information that they do not understand, and transfer knowledge they learn during instruction to help them perform accurately on examinations.

In Experiment 1, participants performed worse on the clicker test in the Delayed condition than in the two Immediate conditions. However, students performed comparably on the posttest and retention test in the three conditions. Analyses also indicated that performance on the clicker questions in the Delayed condition is a better predictor of how well students will perform on subsequent tests than is performance on the clicker questions in the Immediate conditions. Teachers should present clicker questions in the classroom after a delay to optimize the effectiveness of the clicker approach. Thus, teachers will be able to cover or alter learning material based off of clicker responses to maximize the students' overall understanding of material before the midterm and final exam.

Experiment 2 demonstrates that utilizing the clicker during the classroom increases students' retention of the material and enables students to perform better than those who have not studied the material even on questions that require far transfer capabilities. Students in the experimental group were able to learn material and apply it to

more difficult questions, as seen in Transfer 2 and 3 questions, independent of experimental condition.

References

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97-123. doi: 10.1037/h0033773

Anderson, L. S., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2011). Conserving time in the classroom:The clicker technique. *Quarterly Journal of Experimental Psychology, 64,* 1457-1462. doi:10.1080/17470218.2011.593264

Anderson, L. S., Healy, A. F., Kole, J. A., & Bourne, L. E., Jr. (2013). The clicker technique: Cultivating efficient teaching and successful learning. *Applied Cognitive Psychology, 27,* 222-234. doi: 10.1002/acp.2899

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 576-587. doi: 10.1037/0278-7393.6.5.576

Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE-Life Sciences Education, 6,* 9-20. doi: 10.1187/cbe.06-12-0205

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 and Verbal Behavior, 11,* 671-684. doi: 10.1016/S0022-5371(72)80001-X

Craik, F. I., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 12,* 599-607. doi: 10.1016/S0022-5371(73)80039-8

Donovan, W. (2008). An electronic response system and concept tests in general chemistry courses. *The Journal of Computers in Mathematics and Science Teaching, 27,* 369-389.

Guthrie, R. W., & Carlin A. (2004, August). *Waking the dead: Using interactive technology to engage passive listeners in the classroom*. Paper presented at the Tenth Americas Conference on Information Systems, New York.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognitionand recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 562-567. doi: 10.1016/S0022-5371(71)80029-4

Jang, Y. & Huber, E. (2008). Context change in free recall: Recalling long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34, 112-127*. Doi: 10.1037/0278-7393.34.1.112

Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impact fact learning? *Journal of Educational Psychology, 103,* 48-59. doi: 10.1037/a0021977

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Normal (Ed.), *Models of human memory* (pp. 331-373). New York: Academic Press.

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Bulger, M., Campbell, J., Knight, A., & Zhang, H. (2008). 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

Middendorf J., & Kalish A. (1996). The ‘‘change-up’’ in lectures. *National Teaching and*

Learning Forum 5(2). Available online at: http://www.ntlf.com/html/pi/9601/v5n2.pdf (accessed 28 February 2008).

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519-533. doi: 10.1016/S0022-5371(77)80016-9

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science, 323,* 122-124. doi: 10.1126/science.1165919

Smith, R. W., & Healy, A. F. (1998). The time-course of the generation effect. *Memory & Cognition, 26,* 135-142. doi: 10.3758/BF03211376

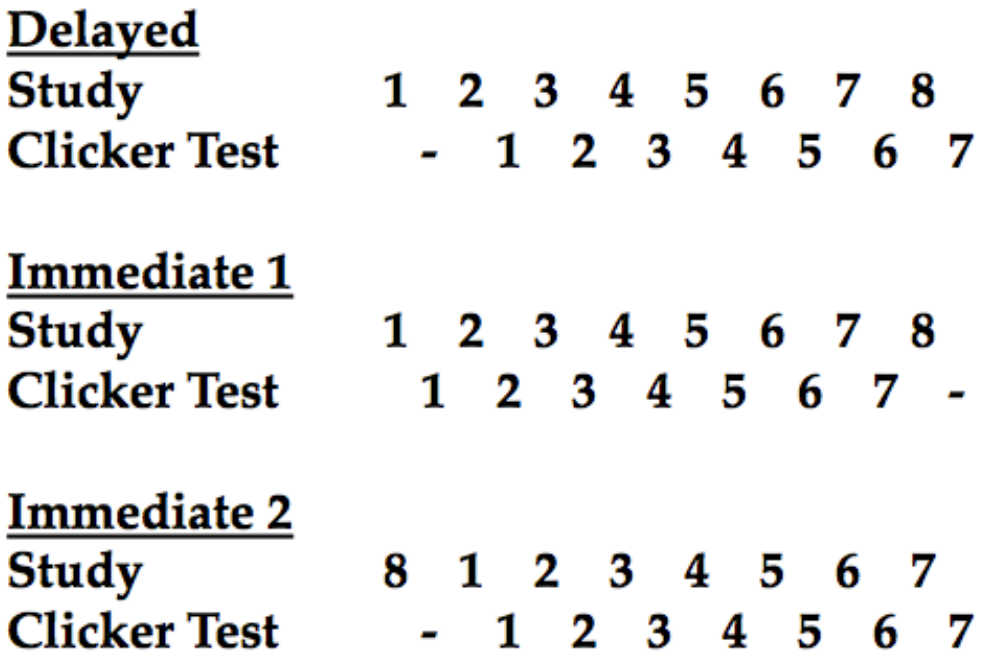

Figure 1. For Experiment 1, temporal orderings of study and test blocks in each experimental condition were presented during the training phase. Each numeral represents a block, including a different category of facts.

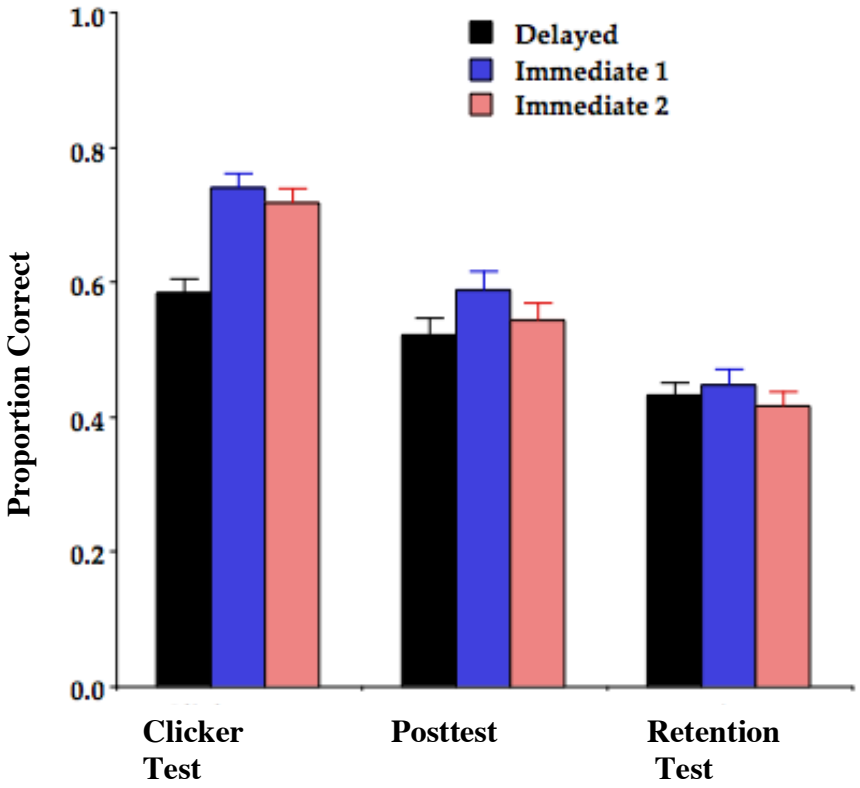

Figure 2. Proportion correct at test as a function of experimental condition and phase.

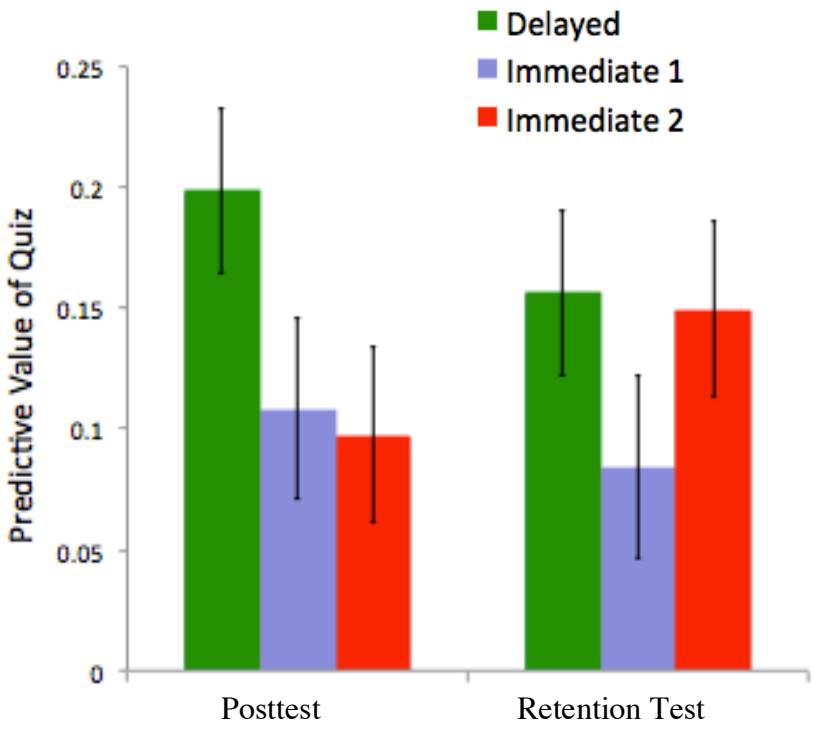

Figure 3. The Delayed condition is a better predictor of performance on posttest and retention test in comparison to both of the Immediate conditions.

Appendix A

Instructions in Experiment 1

In this experiment you will be shown a total of 64 facts about plants. You will be shown 8 facts to learn at a time, and then you will be tested on your ability to remember those facts. Specifically, when you are tested, you will be presented with part of the fact (e.g., A flower that has thorns is the _____?) and four possible answer choices, from which you will select the correct letter answer using the keyboard. The presentation order of the facts will change, so don't rely on that to help you. Instead, associate each fact with each plant. After answering each test question, you will receive feedback indicating the correct answer. This learn/test sequence will continue for about 20 minutes. After a break, you will be tested over all of the facts. Try to remember as many facts about the plants as you can. Again, the presentation order of the facts will change, so don't rely on that to help you. Instead, associate each fact with each plant name. Notes: While being tested, you must select your response within 9 seconds, or the test will automatically advance to the next question. If you do not remember a fact, either guess or press the "Space" key to advance to the next question. If you have any questions please ask the experimenter now. After you have asked your questions, or if you have no questions, press the "Space" key when you are ready to begin.

Appendix B

Experiment 2: Pilot Study

This pilot study involves preliminary tests that examine an initial version of a subset of stimuli used in Experiment 2.

Method

Participants. Ten undergraduate students taking a course in General Psychology at the University of Colorado at Boulder participated in this experiment as a pilot study during the summer session. Due to changes in the pilot study, only the last 3 participants were tested on both a posttest and retention test. The remaining 7 students were tested only on the retention test. Six additional undergraduate students taking a course in General Psychology at the University of Colorado at Boulder took the pretest. These students did not participate in the pilot study.

Materials. A pretest was given to students in the Introduction to Psychology class who participated in experiments outside of this present experiment to gauge the students' baseline understanding of the basics of research methods. The pretest consisted of 28 four-alternative multiple-choice questions that were taken directly from the old and new questionnaires (see the description of these questionnaires below). The first half of the questions was taken from the old questionnaire, whereas the second half of the questions was taken from the new questionnaire. Participants used a pen or pencil to fill out this paper-based pretest.

For the main pilot experiment, which included two parts, every participant was trained and tested on an Apple Mac mini computer with the procedure programmed in MATLAB. The slides used during the study phase were taken directly from the first

lecture of an introductory class on psychological statistics and research methods at the University of Colorado at Boulder. The slides were primarily relevant to the basics of research methods. We also utilized three different quizzes; during Part I each participant was presented with one of the three quizzes, depending on condition. Participants were then tested on both a posttest and a retention test, both of which consisted of 14 multiple-choice questions. An old and new questionnaire, each consisting of 14 questions, was used to create both of the posttests. The old questionnaire consisted of all the questions that were asked during the study phase. The new questionnaire consisted of 14 questions that were used solely during the posttest.

Procedure. Participants were first instructed to read the instructions shown on the initial computer screen. These instructions informed the participants that they would be performing a memory task. In this between-subjects design, participants were in one of the three conditions: *recall-only, recognition-only*, or *recall-recognition*. Part 1, the study phase, consisted of three different sections, and in each section, instructions were first given on a screen informing the participants how many slides they had to study in the specified amount of time. All participants were allowed to study the slides for the same amount of time. In the *recall-only* condition, participants were asked a question and prompted to use the keyboard to type in their answer on the screen. They could then proceed to the next question by hitting "enter," but they were asked to generate some response before continuing. Participants in all conditions received feedback (the correct response) immediately after they entered their answer. In the *recognition-only* condition, participants were to complete a test in the same four-alternative multiple-choice format as the pretest. Participants were given a question with answer options (i.e., "a," "b," "c," and

"d"), and they were prompted to type in the letter corresponding to the correct answer and to hit "enter" to continue. Again, they were required to type in some response before proceeding. In the *recall-recognition* condition, the clicker questions were presented alone, and students were asked to type in their answer choice accordingly before the choices were presented. Then, they answered the question in multiple-choice format. In both instances, they were required to type in some response before proceeding. In Section 1, participants had to study two slides for at least 5 minutes. They could go back and forth between these two slides for 5 minutes; afterwards, the participants would be allowed to proceed to take the quiz, or they could take an extra 2 minutes to study the material further. The first quiz consisted of three questions. Sections 2 and 3 consisted of four slides each with a minimum of 10 minutes and maximum of 14 minutes to study the slides. The quiz after Section 2 consisted of seven questions, whereas the quiz after Section 3 consisted of four questions. The quizzes immediately followed all three sections, and the correct answer was given immediately following each of the participants' responses in all three conditions. After the study phase was complete, the participants took a posttest, which consisted of 14 multiple-choice questions. The questions were selected from a total of 28 questions, 14 of which came directly from the study phase (old questionnaire), and the other 14 of which were new posttest questions (new questionnaire). Questions were counterbalanced across all training conditions. However, feedback was not given during either of the posttests in all three training conditions. In counterbalancing Condition 1 participants received the even-numbered questions that were the same as the quiz during the study phase and the odd-numbered questions from the new list of questions. In counterbalancing Condition 2, participants

received the odd-numbered questions that were the same as the quiz during the study phase and the even-numbered questions from the new list of questions. All questions were then randomized for the order in the posttest and retention test. The first 5 participants had the same randomized question order for their posttest.

Each participant came back exactly 1 day later to be tested for Part II. Participants who were in counterbalancing Condition 1 for the posttest were in counterbalancing Condition 2 for the retention test. Participants who were in counterbalancing Condition 2 for the posttest were in counterbalancing Condition 1 for the retention test.

Results

All 10 of the participants were tested on a retention test. Although there were not many participants in the pilot study, there was a trend for best performance on the retention test by the recognition-only group ($M = .976$, $SEM = .024$), worst performance by the recall-only group ($M = .690$, $SEM = .100$), and intermediate performance by the recall-recognition group ($M = .893$, $SEM = .075$), $F(2,7) = 1.717$, $MSE = 0.076$, $p = .247$. On the retention test, participants had better performance on the new questions ($M = .871$, $SEM = .075$) than on the old questions ($M = .843$, $SEM = .065$). However, this difference was not statistically significant, $F(1,7) < 1$. The interaction of group and question type was also not statistically significant, $F(2,7) < 1$.

Discussion

Participants who were assigned to the recognition-only condition displayed the best performance on the retention test compared to participants in both the recall-only and the recall-recognition conditions. This trend was expected to occur if transfer-appropriate processing overwhelmed any advantage due to retrieval practice. Because all

participants were tested in the recognition condition, it was expected that those participants who were trained in the recognition condition would perform best. Similarly, participants who were trained under drastically different procedures would perform the worst; this difference explains why participants in the recall-only cognition had the worst performance on the retention test. According to transfer-appropriate processing, if professors decide to utilize the clicker technique, then they should test students using the multiple-choice method. The original hypothesis stated that participants in the recall-only condition would perform best due to the greater effort involved with generating a response; recall would facilitate greater retention relative to recognition. Although this was not the case, there were limitations to the pilot experiment that could have affected the results. The results and the statistical nonsignificance may be due to the small sample size. Participants were tested on the retention test exactly 24 hours following the posttest; this delay was increased to 48 hours in Experiment 2 so that examination of long-term memory can be determined more effectively. If participants in the recognition group still perform better after the consideration of these limitations, then it is important to directly test for transfer-appropriate processing by adding a posttest that includes both recognition and recall. These findings will be important for professors who apply the clicker technique in their classroom. If transfer-appropriate processing is the overarching explanation, then professors would want to have clicker questions and exams in the same format. Additionally, participants did extremely well on the pretest (M = 82%) so question difficulty was increased and more questions were generated for the posttest and retention test in Experiment 2.