Digital Commons @ Assumption University

Psychology Department Faculty Works

Psychology Department

2017

Clickers Can Promote Fact Retention But Impede Conceptual Understanding: The Effect of the Interaction Between Clicker Use and Pedagogy on Learning

Amy M. Shapiro
University of Massachusetts at Dartmouth

Judith Sims-Knight
University of Massachusetts at Dartmouth

Grant V. O'Rielly
University of Massachusetts at Dartmouth

Paul Capaldo
University of Massachusetts at Dartmouth

Teal Pedlow
University of Massachusetts at Dartmouth

See next page for additional authors

Follow this and additional works at: https://digitalcommons.assumption.edu/psychology-faculty

Part of the Education Commons, and the Psychology Commons

Recommended Citation

Shapiro, A. M.; Sims-Knight, J.; O'Rielly, G. V. ; Capaldo, P.; Pedlow, T.; Gordon, L.; and Monteiro, K. (2017). Clickers Can Promote Fact Retention But Impede Conceptual Understanding: The Effect of the Interaction Between Clicker Use and Pedagogy on Learning. *Computers & Education* 111: 44-59. https://doi.org/10.1016/j.compedu.2017.03.017

This Article is brought to you for free and open access by the Psychology Department at Digital Commons @ Assumption University. It has been accepted for inclusion in Psychology Department Faculty Works by an authorized administrator of Digital Commons @ Assumption University. For more information, please contact digitalcommons@assumption.edu.

Authors
Amy M. Shapiro, Judith Sims-Knight, Grant V. O'Rielly, Paul Capaldo, Teal Pedlow, Leamarie Gordon, and Kristina Monteiro

This article is available at Digital Commons @ Assumption University: https://digitalcommons.assumption.edu/psychology-faculty/1

Computers & Education 111 (2017) 44–59

Contents lists available at ScienceDirect

Computers & Education

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

Clickers can promote fact retention but impede conceptual understanding: The effect of the interaction between clicker use and pedagogy on learning

Amy M. Shapiro [a, *], Judith Sims-Knight [a], Grant V. O'Rielly [b], Paul Capaldo [c], Teal Pedlow [a], Leamarie Gordon [a, 1], Kristina Monteiro [a, 2]

[a] Psychology Department, UMass Dartmouth, North Dartmouth, MA 02747, United States
[b] Physics Department, UMass Dartmouth, United States
[c] Biology Department, UMass Dartmouth, United States

ARTICLE INFO

Article history:
Received 2 August 2016
Received in revised form 7 January 2017
Accepted 28 March 2017
Available online 31 March 2017

1. Introduction

Personal response systems have received a great deal of attention in the popular media and educational research communities. Most commonly called *clickers*, these devices allow students to send responses to questions posed by the instructor either through a wireless device or a smart phone application. The questions are typically projected to the class on a screen and a receiver collects responses. The instructor is able to project class responses in a graph within seconds. Such systems are widely available on the commercial market and are in widespread use by millions of students, across hundreds of secondary schools, colleges and universities nationwide and internationally. Given the ubiquitous use of the technology, it is important to fully understand its effect on learning, the conditions in which it is beneficial, and best practices for instructors using different pedagogical styles.

Investigations into the educational benefits of clickers have shown that they are effective for imparting factual information during large lecture classes (Lin, Liu, & Chu, 2011; Mayer, Stull, et al., 2009; Shapiro & Gordon, 2012, 2013). While there are reports of clickers being used in classrooms where conceptual learning has improved, these studies typically have used clickers as one of several interventions, such as submitting answers to questions about pre-class readings and small group problem solving workshops (Crouch & Mazur, 2001; Mazur, 2009). Thus, little work has been done to isolate the effect of clickers on conceptual understanding in courses that do not also provide other scaffolds for deep learning. To address this gap, the present study isolated the effect of clickers to the extent possible in live classrooms that used differing pedagogical approaches. The aim was to determine whether clicker-enhanced fact retention and conceptual understanding advantages

* Corresponding author.
E-mail addresses: ashapiro@umassd.edu (A.M. Shapiro), jsimsknight@umassd.edu (J. Sims-Knight), gorielly@umassd.edu (G.V. O'Rielly), pcapaldo@umassd.edu (P. Capaldo), lt.gordon@assumption.edu (L. Gordon), kristina_monteiro@brown.edu (K. Monteiro).
[1] Current affiliation: Psychology Department, Assumption College, 500 Salisbury Street, Worcester, MA 01609, United States.
[2] Current affiliation: Warren Alpert Medical School, Brown University, Providence, RI 02912, United States.

http://dx.doi.org/10.1016/j.compedu.2017.03.017
0360-1315/© 2017 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

can be replicated in classrooms that primarily employ lecture, or a combination of lecture and other instructional strategies designed to impart conceptual understanding.

The following section provides a review of the literature related to clicker-assisted learning. Specifically, it offers some background about clicker effects on learning outcomes, their use in the context of differing pedagogical contexts, and the interaction between clicker use and learner characteristics. Two experiments designed to explore clicker effects in differing pedagogical contexts, and the role of learner characteristics in those effects, are then described.

2. Literature review

2.1. Clicker-assisted learning effects

Published reports indicate clickers are being used for many purposes in classrooms. Among these are teaching case studies (Brickman, 2006; Herried, 2006), replicating published studies in class (Cleary, 2008) and electronic testing (Epstein et al., 2002). Most commonly, though, clickers are used for promoting attendance and participation, and for assessing students' comprehension of class material in real time (Beekes, 2006; Poirier & Feldman, 2007; Shih, Rogers, Hart, Phillis, & Lavoie, 2008).

The literature generally indicates that clickers have positive effects on attendance, class participation, and class enjoyment across various disciplines (Boscardin & Penuel, 2012; Draper & Brown, 2004; Hatch, Jensen, & Moore, 2005; Ribbens, 2007; Stowell & Nelson, 2007; Trees & Jackson, 2007). Some studies have assessed learning outcomes. While a small number of studies report either no or little effect of clicker use on factual knowledge acquisition, methodological issues or factors related to the implementation of the technology in the classroom are often factors in those studies (see Shapiro (2009) for a summary and critique of that literature). The majority of studies examining the effect of clicker use on factual knowledge acquisition have demonstrated a positive effect of the technology.

Ribbens (2007) and Morling, McAuliffe, Cohen, and DiLorenzo (2008) both compared clicker and no-clicker classes and found that students did significantly better on tests when they used clickers to answer questions about the lecture material. Shapiro (2009) targeted specific exam questions with in-class clicker questions and compared performance on targeted test questions with that of a control class that did not use clickers. Performance on clicker-targeted test questions was significantly higher in the clicker class. Shapiro and Gordon (2012) controlled subject and item differences in a similar study. Though reduced in magnitude, the results replicated those of Shapiro (2009). Students performed significantly better on exam questions when they had previously answered clicker questions designed to target the same basic information.

Less work has been reported on the effect of clickers on conceptual understanding in classrooms. Among the most well known interventions using the technology is that of Mazur and colleagues, who report significant gains in student performance on standardized concept inventories in physics (Crouch & Mazur, 2001) when using a peer instruction method that involves the use of clickers. Indeed, Crouch and Mazur report a doubling or tripling of pre-post test gains on standardized learning assessments (the Force Concept Inventory and the Mechanics Baseline Test) as compared to semesters in which the course was taught using traditional instructional methods. Of course, it cannot be determined whether these improvements are due to the peer activity or to the clickers.

Levesque (2011) also explored conceptual understanding in an introductory genetics course. She targeted the content from eight class periods (the content on a single exam) and examined the effect of clicker use on exam performance. The study compared classes that used clickers to control classes in other semesters on the same content. In all semesters students listened to lectures and completed problems for homework assignments. In the clicker semesters, lecture periods included individual and group problem solving, with clickers used to submit responses. In the control semesters the instructor worked out the problems, with students volunteering to participate by raising their hands if they chose to. The exam questions covered the same material as the clicker questions, though they were not identical.

The investigation's major findings were (a) a correlation between the percentage of clicker questions answered in class and exam performance, and (b) a correlation between attendance and exam performance in the clicker semesters, but not in the no-clicker semesters. The author argues on this basis that the clicker questions were effective in promoting problem solving ability. While the article does not provide parametric analysis of exam performance between clicker classes and no-clicker classes, the difference in the reported performance means between control and experimental classes is rather small. The clicker classes scored means of 75.3 and 77.8, respectively on the exam (for a rough mean of 76.5). The no-clicker classes scored means of 77.6 and 72.5, respectively (for a rough mean of 75.1). Thus, it does not appear the addition of a relatively small number of in-class clicker problems to the assigned homework problems made a meaningful difference (Levesque, 2011).

2.2. Clicker use in context

The evidence for the effect of clicker questions on factual knowledge gains is fairly compelling, though the evidence for improvement of deeper, more conceptual understanding is less clear. One goal of the present investigation was to conduct a highly controlled study of the effect of clicker use on both factual and conceptual knowledge. We sought to do so mindful of the broader classroom context, which is a factor not often addressed in the clicker literature. By context, we mean the overall pedagogical approach used by the instructor, assignments or activities required outside of class meetings, and the instructor's learning goals for students.

A very useful bibliography of clicker studies is maintained by Derek Bruff at Vanderbilt University ("Classroom Response System ("Clickers") Bibliography," 2016). The database provides citations for clicker studies in the humanities, social sciences, natural sciences and engineering. Based on our reading of the literature, there do not appear to be clear, systematic differences in clicker effects between disciplines. We suspect the pedagogy employed within a course, regardless of domain, is likely to be more relevant to clicker effects than the domain. The role of other pedagogical practices employed in the classroom is seldom addressed or introduced as a variable in clicker studies, however, in spite of the fact that practices can vary widely. Some instructors take a "traditional," didactic approach to teaching. This approach reflects a "bottom-up" philosophy of learning, which builds toward conceptual understanding and problem solving only after the learner has acquired a foundation of knowledge (e.g., Bloom's Taxonomy; see Hill & McGraw, 1981; and Krathwohl, 2002). Others take a more constructivist approach. In these classrooms, the amount of didactic lecture varies, but students are required to spend considerable effort either inside or outside of class solving novel problems that require application and expansion of existing knowledge and/or skills (Barrows, 1996; Blumenfeld et al., 1991; English & Kitsantas, 2013; Gormally, Brickman, Hallar, & Armstrong, 2009; Jonassen & Hung, 2008; Linn, Songer, & Eylon, 1996; Savery, 2006; Schmidt, Loyens, van Gog, & Paas, 2007; Strobel & van Barneveld, 2009; Wirkala & Kuhn, 2011). Using this approach, information can be acquired through lecture and readings, but much of students' deeper understanding is acquired as students engage with the material by solving novel problems, doing experiments, or working on projects (Eysink et al., 2009).

In recent years, college educators have increasingly begun to consider or implement more interactive forms of instruction (Lambert, 2012; Rimer, 2009). As more educators seek to move away from the traditional lecture-based model, the question of classroom context as a variable in studies of learning has become increasingly important. From the standpoint of application and practice, it is important to determine whether the effect of clickers on fact retention and conceptual understanding differs between classrooms adopting a didactic as opposed to an active learning approach.

A small number of researchers have noted the importance of context to their reported results. Mazur (2009), in particular, has written about the prevalence of pedagogical practices apart from clicker use in his courses. Nonetheless, it is often overlooked that his method relies very little on clickers *per se*. As Mazur explains, "I often meet people who tell me they have implemented this "clicker method" in their classes, viewing my approach as simply a technological innovation. However, it is not the technology but the pedagogy that matters (2009, p. 51)." Students in Mazur's physics courses were required to do readings designed as pre-lecture material and to answer questions about those readings before arriving in class. In addition, they spent entire classes, called 'workshops', watching the instructor work through a problem, then working on multiple problems in small groups with the instructor merely circulating and offering support. There were also additional homework assignments, in which more problems were assigned. Thus, students were problem-solving or otherwise actively engaged with the material, either with the instructor or one another, most of the time. Even when reading, the assignment was designed to be active, as the students were required to submit answers to questions before coming to class. Clickers were used during lecture time (which was much reduced as compared to the traditional lecture-based classroom), and combined with peer interaction, to give students opportunity to engage with the material offered in lecture, and to practice problem solving. Mazur's use of such approaches to promote deep learning and critical thinking is in step with a rich literature on problem-based inquiry and critical thinking skill development (see reviews by Abrami et al., 2015; Wilder, 2015).

Levesque (2011) has also noted the importance of activities outside of clicker use *per se*, on students' ultimate success. She points specifically to the value of students completing assigned homework problems. In reflecting on the correlation she found between in-class clicker use and exam performance, she suggested that clicker use might be effective, in part, because "the involvement of students in solving problems during class increased the likelihood that they would attempt to complete practice problems outside of class. In this scenario, by being involved in problem solving in class, a student would gain the confidence to attempt problems independently outside of class, which would be reflected in exam performance (p. 416)." Levesque asserts, then, that the clicker questions themselves did not directly affect performance. Rather, she suggests they gave students courage to tackle more of the practice problems outside of class, and that the heavy practice outside of class was a likely source of benefit for her students.

2.3. The relationship between the effect of clickers and student variables

Clicker use may be differentially effective for different kinds of students. Clickers may not particularly help students who possess attitudes and skills that effectively promote learning. Because those students will tend to engage in activities that lead to deep understanding and retention of material, it is likely they will do equally well with or without clickers. Contrarily, we would expect students who had attitudes and learning habits that are less effective for learning to be helped by clickers.

Among potential positive attitudes and skills that may affect clicker-assisted learning, we chose three that have been well documented. *Prior knowledge* is one such predictor of positive learning outcomes, most likely because a foundation of existing knowledge provides a foundation and scaffold upon which to build new knowledge (e.g., Kalyuga, 2012; Shapiro, 1999; Williams & Lombrozo, 2013). Second, various measures of *self-regulated learning* and metacognition have been shown to be associated with positive learning outcomes (e.g., Azevedo, 2005, 2015; Greene, 2015; Pintrich & Schrauben, 1992, pp. 149–183). Third, *deep learning strategies*, one of the subscales of the Revised Study Process Questionnaire (SPQ, Biggs, Kember, & Leung, 2001), has been found to be positively associated with academic achievement (c.f. meta-analysis by Watkins, 2001).

It was chosen as a variable of interested here because the relationship between deep learning strategies and academic achievement is affected by achievement-related classroom behaviors (Choy, O'Grady, & Rotgans, 2012), and has been found to change as a function of educational innovations (Kember, Charlesworth, Davies, McKay, & Stott, 1997). While deep and shallow learning strategies have shown to be separate measures, rather than opposite extremes on a single scale, shallow learning also has been shown to be related to achievement. Specifically, the second subscale of the SPQ, shallow learning, has been found to be negatively related to academic achievement (Watkins, 2001).

3. Experiment 1

The experiment was designed to test the effect of clickers on comprehension and learning of in-class material in a course using traditional, lecture-based pedagogy. It was designed to evaluate memory for rote, factual material as well as comprehension and retention of deeper, more conceptual learning. We make three hypotheses based on the existing clicker-assisted learning literature. First, we expect to replicate prior work that has demonstrated factual and conceptual clicker questions will boost performance on factual exam questions, both because the introduction of these questions in class affect the time spent working with the material and invokes the testing effect. Second, we predict conceptual clicker questions will boost conceptual exam question performance, because it will increase time spent with the material. Constructivist theory also predicts such an effect, as the questions provide practice thinking creatively about the material, thus deepening understanding. Third, we predict student variables such as prior knowledge and students' approach to learning will mediate the effect of clicker questions on learning outcomes.

The research was conducted in a live classroom in order to maximize the ecological validity of the findings. The work was conducted in an introductory course in biology. All lectures were accompanied by PowerPoint presentations that projected main points and illustrations onto a large screen. In-class clicker questions were integrated into the PowerPoint presentations, with individual slides dedicated to single questions. Students were required to purchase their clickers for each class. The experiment included four treatment conditions (factual clicker, conceptual clicker, enhanced control and simple control) in a repeated measures design. Performance on targeted exam questions was used as the dependent variable, with performance on both factual and conceptual exam items assessed in all conditions. Factual exam questions probed knowledge of basic facts (such as a definition, research finding, or formula) presented in class, while conceptual questions required students to go beyond rote memory and make connections between the underlying concept and other material in the course, or apply the concept to a problem or real-life scenarios.

3.1. Method

3.1.1. Participants and the course

Participants were all undergraduate students at a university in the Northeast United States, enrolled in Introduction to Biology (an introduction to cells, genetics and evolution). A total of 858 students were enrolled (54% female) across semesters, and had a mean grade point average (GPA) of 3.05/4.0. The average age was 19.03 years. Students were biology nonmajors taking the course to fill a distribution requirement, and were primarily in their first year. The participants largely reflected the population of the campus, which is 86% Caucasian, 5% African American and 3% Hispanic.

The experiment ran over four consecutive semesters beginning in fall 2010, with minor variation in enrollments in each semester. The experiment received IRB approval each year of the study as an exempt project. Students were informed about the research activity in their classes through an IRB-approved consent statement in the syllabus, and they were able to decline participation with no penalty. A small amount of extra credit was offered in each course for completing a survey about learning habits and behaviors, and students were offered the opportunity to earn the extra credit through an alternative assignment should they choose to opt out of the study. No students declined participation.

The instructor had 11 years of university teaching experience at the start of the project and received excellent evaluations by his department and students in prior semesters. He had been using clickers in his classroom for several years prior to the study.

3.1.2. Experimental conditions and design

Four conditions were developed to test the study's hypotheses. Two of the conditions directly tested clicker effects, as they involved presenting students with clicker questions designed to probe information needed to answer specific exam questions. The other two conditions were controls. The *simple control* was a simple “do nothing” control condition, while the *enhanced control* condition directed students' attention to the importance of the relevant material while refraining from offering clicker questions. The enhanced control condition was included to test the hypothesis that clicker questions raise test scores by drawing students' attention to information more likely to appear on exams, thus increasing study of the relevant material. The four conditions were as follows:

3.1.2.1. Factual clicker. The factual clicker condition was included as a direct test of the effect of factual clicker questions on memory and understanding of class material. A PowerPoint slide containing a factual question about a lecture topic was presented after the instructor conveyed the information and any student questions were asked and answered. Factual questions required students to demonstrate memory for definitions, formulas, concepts or other facts presented in class. Factual questions required no more than rote memorization of a definition, fact or formula to answer. If a question required merely that numbers be applied to a formula, it was considered factual.

3.1.2.2. Conceptual clicker. The conceptual clicker condition was included as a direct test of the effect of conceptual clicker questions on memory and understanding of class material. A PowerPoint slide containing a conceptual question about a lecture topic was presented after the instructor has conveyed the information and any student questions were asked and answered. Conceptual questions were designed to test deeper understanding of the material, such as the relationship between the underlying concept and other material in the course or the application of the information to a problem or real-life scenarios. They often rely on a student's ability to use factual knowledge (e.g., definitions, facts or formulas) to address a novel situation or problem, thus demonstrating some depth to students' understanding of the meaning or broader impact of a course topic. In other words, understanding beyond rote memorization of material from the book or class lecture should be necessary to answer these questions.

3.1.2.3. Enhanced control. The enhanced control condition is one of two control conditions included in the study. It was intended to illuminate whether factual clicker questions directly affect what students learn in class or if clicker effects may be attributed at least in part to some class information being highlighted more than others. When the information necessary to answer an exam question was targeted as a “flagged” or enhanced item, it was presented in class and highlighted on the projected slide. This was done by using the instructor's remote to change the font color (usually to red or bright yellow) and pulse the text. In addition, the instructor announced, “This information is very important. It is likely to be on your test.” Flags were variably offered before, during or after the presentation of the relevant information. Comparing performance on items when they were used in the enhanced control condition with performance when they are used in the factual clicker and simple control conditions tested whether factual clicker questions provide greater benefit than directing students to important information, and promoting additional study.

3.1.2.4. Simple control. The simple control condition provided a baseline measure of student performance when no intervention was offered. When an item was included in this condition the information was discussed or presented in class just as it was when the item is included in the clicker conditions, except that no clicker question relevant to that information was offered in class. No special emphasis was put on material when included in the simple control condition. Comparing performance on items when they were used in the simple control condition with performance when they are used in the factual and conceptual clicker conditions provided a direct test of the effect of clickers in the classroom.

The dependent variable was performance on conceptual and factual exam questions. There were roughly equal numbers of each type of exam question embedded within each exam. Factual exam questions were written to require students to demonstrate they knew the rote, factual information presented in class. These included definitions, theories, formulas and other basic concepts. Conceptual exam questions were designed to test deeper understanding of the material, such as the relationship between the underlying concept and other material in the course or the application of the information to a problem or real-life scenario.

Each exam question was assigned to a different experimental condition in each of the four semesters of the study, creating a repeated measures design. This procedure allowed performance on each experimental exam question to be compared when targeted by a different treatment. Otherwise put, this design allowed the effects of the four conditions to be compared without the confounding variable of item differences affecting the results. The use of items from each of the four conditions rotating between semesters also allowed subjects to participate in all four conditions of the study. This element of the experimental design created a within-subject design that also removed the potential confounding variable of subject differences between conditions. To clarify the experimental design, a hypothetical illustration of item assignment in a course is provided in Table 1.

Items were counterbalanced among conditions within each semester to equalize the distribution of factual and conceptual items sets across the semester to the extent possible. The number of items in each condition was also balanced to the extent possible for each exam. There were three exams given in each course every semester. To achieve equal distribution of factual and conceptual item sets in each condition per semester, and the number of item sets per condition on each exam, the questions were first segmented into groups by the exams in which they would be tested. They were then separated by whether they were factual or conceptual exam questions so that a fairly even number of these could be distributed across experimental conditions within each exam. Assignment to each condition was then assigned to “blocks” A, B, C, or D in round-robin style, in the order the topics would appear in the class. This was done to counterbalance the temporal presentation of the information. In other words, items covered early versus late in each exam unit were spread evenly across conditions.

As part of the item set validation procedure (to be discussed in section 2.1.3.2), content experts were asked to rate how strongly material in one item set related to the others. These ratings were obtained to allow control of possible “spillover effects” between item sets and prevent cross contamination between conditions. That is, it was important to prevent clicker questions for one item set from affecting performance on exam questions assigned to a different experimental condition. For this reason, the counterbalancing procedure described here was modified for items that did not receive very low ratings (3 or

Table 1
Hypothetical assignment of targeted exam question items in one course to each experimental condition over four semesters. Each subject participates in each condition within a semester, and each targeted exam question serves in each condition over the four semesters.

Semester and Subject Numbers	Targeted Exam Questions in Each Experimental Condition
Semester 1	Factual Clicker: 1-8
Subjects 1-100	Conceptual Clicker: 9-16
	Enhanced Control: 17-24
	Simple Control: 25-32
Semester 2	Simple Control: 1-8
Subjects 101-200	Factual Clicker: 9-16
	Conceptual Clicker: 17-24
	Enhanced Control: 25-32
Semester 3	Enhanced Control: 1-8
Subjects 201-300	Simple Control: 9-16
	Factual Clicker: 17-24
	Conceptual Clicker: 25-32
Semester 4	Conceptual Clicker: 1-8
Subjects 301-400	Enhanced Control: 9-16
	Simple Control: 17-24
	Factual Clicker: 25-32

lower on a 7-point scale) for possible spillover effects. In those cases, items were grouped together into the same condition to keep effects of each treatment isolated from the others. In this way any spillover from a clicker question to an unintended exam question was retained within a condition rather than contaminating another condition. Many of the items in the physics class were grouped by content as a result of this procedure.

3.1.3. Materials and equipment

3.1.3.1. Hardware. The instructor used the Interwrite PRS system, produced by the eInstruction Company. The student remote had 5 response buttons labeled A-F plus ten numeric response buttons. It also had an LED display that displayed a message to inform students their response has been recorded, as well as an on-off button. The system required the instructor to attach a receiver to his computer that recorded responses and stored them in data files.

3.1.3.2. Item set development and validation. Development and validation of item sets (sets of clicker questions, accompanying lecture slides, and paired exam questions) followed a highly structured process. A sample item set from the course is provided in Fig. 1.

To establish validity of the experimental materials, a detailed validation procedure was developed and carried out prior to beginning of the experiment, in which two faculty members from the Biology Department were recruited as content experts to assess the materials in their respective fields. Interrater reliability was attained to assure that the factual questions were factual and the conceptual questions were conceptual. They also noted and helped correct any ambiguities or problems with each question. To rule out the possibility that clicker questions affected performance on exam questions they were not intended to target (‘spillover’ effects), raters were also asked to determine the degree of relationship between each of the exam and clicker questions, and the critical information targeted by each. Finally, the validation procedure also established the quality of the critical slides and the relationships between the critical slides and the exam and clicker questions.

We considered an item acceptable if it met a set of predetermined ratings criteria. A question was considered acceptable for use in the course if (a) the mean rating among raters was 5 on the factual/conceptual scale, (b) each rater gave a rating between moderate and high (4–7), (c) the difference between rater judgments was 2 or less (the mean difference was 0.3), (d) the item was rated a mean between 4 (good) and 7 (excellent) for overall quality, and (e) if the mean relationship rating between the question and the target information was between 5 and 7. Questions that did not meet these criteria were revised by the instructor and sent back to the raters for re-scoring. Items that failed to meet the criteria after 3 rounds of this iterative process were excluded from the study. The majority met standards after 2 rounds. The mean difference between the judges’ factual and conceptual ratings was less than 0.5.

3.1.3.3. Student learning survey. To gain information about participants' demographic information and approaches to studying and learning that might act as mediating variables, we generated a Student Learning Survey. The first page asked information about students’ age, year in school, gender, GPA and familiarity with course content prior to taking the course.

These items were followed by Biggs et al. (2001) R-SPQ-2F questionnaire; twenty questions about learning motives and strategies. This instrument is a shortened version of the validated Study Process Questionnaire (SPQ; Biggs, 1987). The aim of the R-SPQ-2F questionnaire is to evaluate four specific dimensions of student learning: Deep Motive (DM), Deep Strategy (DS), Surface Motive (SM) and Surface Strategy (SS). Students assessed as having deep motives can be characterized as being

(a) Sample conceptual exam item set in the didactic (biology) course

Target Content
A species is defined as a group/population of animals which can interbreed and produce fertile offspring.

Target Exam Question (Conceptual)
The "Liger" is the hybrid offspring of a lion and tiger. However, the Liger is sterile. What does this tell you about the lion and tiger parents?

A. They share the same genotype
B. They share the same phenotype
C. They are separate species *
D. They live in the same geographic area.

Factual Clicker Question
A group or population which can successfully interbreed and produce fertile offspring is called a:

A. genus
B. phylum
C. order
D. Species*

Conceptual Clicker Question
A horse and a donkey can and will breed together. Therefore, they must be a species.

A. True
B. False *

Content Slide (text highlights appeared only in the enhanced control condition)

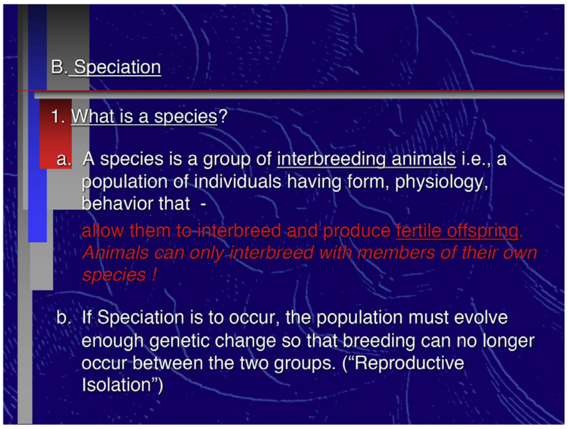


Fig. 1. A sample conceptual exam item set in experiment 1 (a), and a sample factual exam item set in experiment 2 (b), including the critical information slides. Other information slides relevant to each item set were shown in class, though the slide containing the information most directly relevant to the target exam question was considered the critical slide. The correct answer to each question is indicated with an asterisk.

motivated to learn and understand content. Those with deep strategies use learning strategies that are directed at achieving understanding and mastery of the material, such as formulating questions prior to class, resolving areas of misunderstanding after class, re-writing notes, and so forth. Students who are focused on surface learning are more motivated to score well on exams and assignments rather than arriving at a deep understanding of the material. The efforts of surface strategy learners are generally spent engaged in activities such as memorizing definitions and doing as little work as possible to get a good grade.

The Cronbach alpha values for each subscale of the R-SPQ-2F are as follows: DM (0.62); DS (0.63); SM (0.72); and SS (0.57). In administering the questionnaire, we did rearrange the order of the questions. The original order had the items arranged in a pattern, with items from each subscale presented in every fourth position. To avoid response bias that may stem from students unconsciously detecting the pattern, we re-arranged the order of the items so that they were presented in a random order.

Next we included two subscales from the MSLQ (Garcia & Pintrich, 1996; Pintrich, Smith, Garcia, & Mckeachie, 1993), to determine whether students were active or passive learners. The rehearsal subscale determines whether students engage in active strategies aimed at memorization or other surface learning. The organization subscale measures whether students engage in active strategies designed to arrive at a deeper understanding of the material. Garcia and Pintrich (1996) report Crohnbach alpha scores of 0.69 and 0.64 for the rehearsal and organization subscales.

(b) Sample factual exam item set in the problem-oriented (physics) course

Target Content

Conservation of energy

Target Exam Question (Factual)

A skier at rest at the top of a ski run (assumed to be frictionless) has a potential energy of 4.7 x 10^5 J. At some point along the ski run the skier's kinetic energy is 2.3 x 10^5 J. What is the skier's potential energy at this point?

A. 2.3 x 10^5 J
B. 2.4 x 10^5 J (*)
C. 3.5 x 10^5 J
D. 4.7 x 10^5 J
E. 7.0 x 10^5 J

Factual Clicker Question

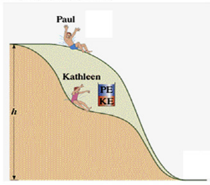
A box sliding along a frictionless surface has a kinetic energy of 300 J when it reaches a ramp (also frictionless). As to block slides up the ramp, when its potential energy is 120 J what will its kinetic energy be?

A. 2.3 x 10^5 J
B. 2.4 x 10^5 J (*)
C. 3.5 x 10^5 J
D. 4.7 x 10^5 J
E. 7.0 x 10^5 J

Conceptual Clicker Question

Tom and Kathleen start from rest at the top of two waterslides with different shapes. Who will have the greater speed when they get to the bottom?

A. Tom will have the greater speed
B. Kathleen will have the greater speed
C. They will have the same speed (*)
D. It depends on the initial height of the waterslides.



Content Slide (text highlights appeared only in the enhanced control condition)

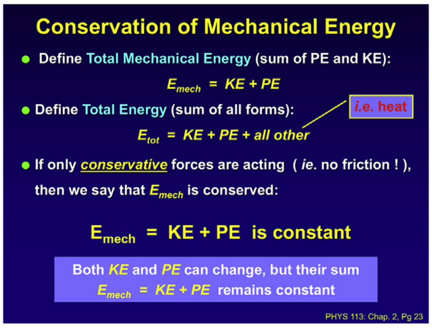


Fig. 1. (*continued*).

Because students often mistake familiarity for knowing or overestimate their level of understanding (Koriat & Bjork, 2005; Kornell & Bjork, 2009; Metcalfe, 1998), we also included a measure of metacognitive self-regulation. We thought this was important because learning effects of clicker questions may stem from the information they provide students about what they don't know or understand about the material. Students who have low self-knowledge may benefit from exposure to evidence about their knowledge level, as provided by clicker questions and their subsequent feedback. Four items from Pintrich's (Garcia & Pintrich, 1996) MSLQ metacognition subscale and four items from Schraw and Dennison's (1994) Metacognitive Awareness Inventory (MAI) were combined to comprise a new measure of self-knowledge. The MSLQ metacognitive self-regulation subscale is a validated scale of metacognition (Cronbach's alpha = 0.79). The items directed at assessing self-knowledge specifically were used for the present study.

3.1.4. Fidelity to the intervention

Several precautions were taken to ensure the instructor was faithful to the study's design and remembered to incorporate the targeted materials in the correct conditions each semester. First, the instructor was provided a packet of *item fidelity forms* at the beginning of each semester, which was used to log the date and condition of each item set during the semester, and the placement of the test questions on the exams. In addition to the instructor retaining records of the intervention in his course, the experimenter and assistants conducted regular fidelity visits during each semester. The class was visited at least once every 2 weeks during the semester to monitor class content for fidelity to the intervention.

3.1.5. Procedure

The instructor created elaborate PowerPoint presentations that offered definitions, bullet points, illustrations, and so forth to augment the verbal content of his lectures. When an item was assigned to the factual or conceptual clicker conditions, the basic critical slide for that information was shown, embedded in the other slides for that day's lecture. At the appropriate time, the targeted clicker question (factual or conceptual) was presented during class for that item. When assigned to the simple control condition, the basic information slide for a targeted item was shown in class in an identical way as the clicker conditions but no clicker questions were presented on that topic. For the enhanced control condition, the enhanced slide was shown in place of the standard slide. In addition, the instructor verbally told the students that the information was very important and would be on the next test. No clicker questions were asked about that information in class. Table 2 summarizes the classroom treatment in each condition.

From the students' point of view, they attended class each day and were required to bring their clickers with them. At unpredictable points during class, the professor would pose a clicker question. Some of the questions were factual and some were conceptual, though the professor did not distinguish between those explicitly. Students were required to answer, the class polling results were presented in the form of a bar graph and the correct answer was eventually revealed. Clicker questions that were related to the study were not distinguished from any others that were not related to the study (though they were all carefully tested beforehand for cross-contamination by the instructors and content experts). Some days no clicker questions were asked and on other days there were multiple questions.

The typical presentation of experimental clicker items occurred after a concept was presented and the instructor invited questions. After all clarifications were made, the clicker question was presented. A bar graph showing the percent of students responding with each choice was then projected for all to see. The correct answer was indicated by the instructor and the reason for the correct answer was discussed if there was any dispute or if the class did not score well as a group (over 90%).

The experimental exam questions were spread roughly evenly across exams during the semester in each course. They were also placed randomly within the exams, and were not identifiable as targeted items to students. The demographic and learning surveys were given after the first exam but before the second so that students would be able to provide accurate information on their study strategies.

3.2. Results and discussion

The data were edited to reduce error in the analysis. Students who answered fewer than 75% of the in-class clicker questions were removed from all conditions of the analysis, as they were not sufficiently exposed to the treatment conditions to be part of the study. Also, students who did not attend at least two exams were not included in the analysis, as they did not provide sufficient data for the within-subject analysis. T-tests between included and excluded subjects in both courses revealed no significant differences in student age, year in college, GPA or prior knowledge between groups ($p > 0.05$ in all cases).

Our first analysis tests hypothesis 1, that both types of clicker questions will boost factual knowledge retention. A repeated measures ANOVA (with Greenhouse-Geisser correction) showed a significant effect of treatment condition on the factual exam questions, $F(2.9, 1272.1) = 26.1$, $p < 0.001$, $\eta_p^2 = 0.06$. Thus, there was an effect of the experimental conditions on factual exam performance. The means (and standard deviations) for each condition are provided in Table 3.

A series of paired *t*-tests revealed the source of differences between conditions on factual exam question performance. The results of the *t*-tests are presented in Table 4, which shows that factual clicker questions (*mean* = *76.8*), conceptual clicker questions (*mean* = 69.8) and the enhanced control condition (*mean* = 78.4) were effective means of boosting retention for factual material over the control group (mean = 64.9). The differences between the factual and conceptual clicker conditions, and between conceptual and enhanced control conditions were significant. The difference between the factual clicker and enhanced control conditions was not significant.

Hypothesis 2 predicted that conceptual clicker questions will boost conceptual exam performance. A repeated measures ANOVA (with Greenhouse-Geisser correction) showed a significant effect of treatment condition on the conceptual exam questions, $F(2.89, 1230.9) = 6.9$, $p < 0.001$, $\eta_p^2 = 0.02$. Despite the significant result, none of the conditions improved performance over the simple control condition. As shown in Table 4, the significant finding was due to superior performance in

Table 2
Treatment of targeted information in each experimental condition.

Experimental condition	Critical slide	Clicker question	Target exam question on exam
Factual Clicker	Plain	Factual	Yes
Conceptual Clicker	Plain	Conceptual	Yes
Enhanced Control	Flagged	None	Yes
Simple Control	Plain	None	Yes

Table 3
Means (and standard deviations) for each condition in experiments 1 and 2.

Experimental Condition				
Experiment/Exam Question Type	Conceptual Clicker	Factual Clicker	Enhanced Control	Simple Control
EXPERIMENT 1 (didactic)				
Factual	69.8 (28.2)	76.8 (26.1)	78.4(25.0)	64.9 (27.3)
Conceptual	62.4 (30.8)	62.3 (33.9)	69.9(27.1)	65.1 (27.3)
EXPERIMENT 2 (problem-oriented)				
Factual	72.8 (24.4)	67.8 (22.0)	70.0(19.5)	68.7 (24.0)
Conceptual	72.1 (27.2)	59.7 (40.9)	51.7(39.8)	68.7 (24.0)

Table 4
T-test results (and df) between conditions for the conceptual and factual exam questions in experiment 1, and conceptual exam questions in experiment 2.

T-test Comparison Pairs						
Course/Exam Question Type	Conceptual Clicker and Factual Clicker	Conceptual Clicker and Simple Control	Conceptual Clicker and Enhanced Control	Factual Clicker and Simple Control	Factual Clicker and Enhanced Control	Enhanced Control and Simple Control
EXPERIMENT 1						
Factual	**-3.51	**3.00	**-5.16	**8.30	0.58	**7.60
	(457)	(455)	(449)	(497)	(494)	(505)
Conceptual	0.01	−1.01	**-3.56	−0.95	**-2.90	1.60
	(430)	(454)	(455)	(447)	(448)	(507)
EXPERIMENT 2						
Conceptual	**2.74	1.22	**5.36	*-2.25	1.67	**-4.24
	(133)	(133)	(133)	(134)	(133)	(133)

*$p < 0.05$; **$p < 0.01$.

the enhanced control condition (mean = 69.9), where students performed better than in the conceptual clicker (mean = 62.4) and factual clicker (mean = 62.3) conditions. None of these conditions boosted performance over the simple control condition, which fails to support our prediction of improved performance on conceptual exam questions in the clicker conditions.

To explore our third hypothesis, that subject variables will be a factor in learning outcomes, we conducted a series of secondary analyses to determine the interaction between subject variables and the independent variables. To conduct these analyses, we divided the subjects at the median score for each subject variable into 'high' and 'low' groups, and treated each subject variable as a 2-level independent variable, to conduct a series of 2 (subject variable: high, low) X 4 (experimental condition: factual clicker, conceptual clicker, enhanced control, simple control) ANOVAs). There was an interaction with deep learning strategies and motivation. As Fig. 2 shows, in the simple control, deep learners performed better than non-deep learners, as the literature on self-regulated learning suggests they tend to do (Azevedo, 2005; 2015). The deep learning advantage is not present in the other conditions, however. The clicker and enhanced control conditions facilitated the non-

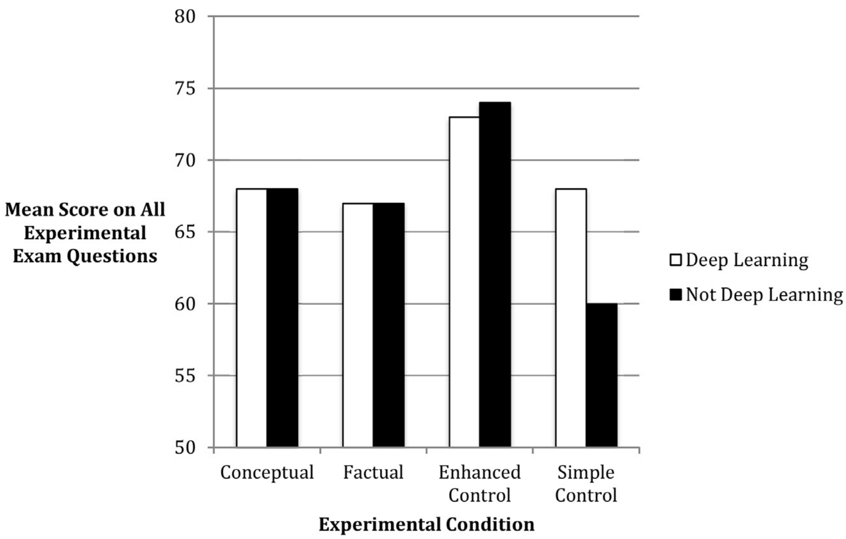


Fig. 2. Secondary analysis of deep learning strategy and experimental conditions in experiment 1.

deep learners. Indeed, when in those conditions the non-deep learners' performance was raised to the level of deep learners, as the difference between the groups was reduced to non-significant levels. This result suggests that clicker questions and the added attention promoted by the enhanced control condition boosted performance for students who do not use deep strategies to the level of their deep-strategy peers. Analyses of the other student variables showed no significant interactions between the experimental conditions and students' metacognitive self-regulation, active learning, shallow learning strategies and motivation, GPA or prior knowledge, $p > 0.05$ in all cases.

In sum, the results support our first hypothesis and replicate prior studies that report factual and conceptual clicker questions increase factual knowledge retention over a simple control condition. Our second hypothesis, that conceptual clicker questions will enhance conceptual exam question performance, was not supported. Our third hypothesis, that student variables would mediate clicker effects, was partially supported by the analysis of deep learning strategies. Specifically, we found that clicker questions brought overall exam performance of students who do not employ deep learning strategies to the level of their deep strategy-using peers. Thus, we have shown that clicker questions have differential effects depending on students' learning orientation.

4. Experiment 2

Experiment 1 replicated much of the prior work on classroom clickers, and provided confirmation of two of our hypotheses in a didactic course. In experiment 2, we were interested in the effect of clicker use in courses that use a different pedagogical orientation, to determine if the same results could be replicated in a less didactic course. Thus, we apply here the same test of clicker effects to a course taught by an instructor who used a problem-oriented pedagogy. We predicted that clicker questions would produce a main effect on factual exam questions, as the testing effect should still apply in this case. We did not predict a main effect of clicker questions on conceptual exam questions. We reasoned that the amount of additional time spent with the material via the clicker questions should be too small in relation to their other active learning activities to have a measurable effect. We did predict, however, that stronger, more knowledgeable students would score differently from their less prepared counterparts, in response to the clicker intervention.

The experimental design, material development process, validation and fidelity procedures, student surveys, and experimental procedures were generally the same as experiment 1. Any alterations are explained in the following section.

4.1. Method

4.1.1. Participants and the course

Participants were all undergraduate students a university in the Northeast United States. Each student was enrolled in Classical Physics I (an introduction to forces, kinematics, energy and momentum) to fill a major requirement and was in the first or second year. A total of 299 students were enrolled (23% female) across semesters, with a mean GPA of 2.93/4.0. The average age was 20.12 years. The participants largely reflected the population of the campus, which is 86% Caucasian, 5% African American and 3% Hispanic. The instructor had 10 years of university teaching experience at the start of the project and received excellent evaluations by his departments and students in prior semesters. He had been using clickers in their classrooms for several years prior to the study.

4.1.2. Materials and equipment

4.1.2.1. Hardware. The instructor adopted the ResponseCard RF system produced by Turning Technologies. This remote came equipped with 10 buttons that could be used interchangeably for alphabetic (A-J) or numeric (0–9) input. The system required the instructor to attach a receiver to his computers that recorded responses and stored them in data files.

4.1.2.2. Item set development and validation. Development and validation of item sets followed the same general procedure as in experiment 1. The process garnered a total 38 item sets, and all were used in the study, with one alteration. Although the problem set items in experiment 2 met the same evaluation criteria as the experiment 1 items for overall quality and the relationship between the questions and target information, the physics experts had more difficulty agreeing about the factual and conceptual designation of each item. For this reason, we brought in a fourth rater and obtained agreement by at least 3 of 4 physicists (three raters and the instructor) for each item used in the experiment. The raters' initial disagreement and its resolution is described in Appendix A. A sample item set is provided in Fig. 1.

4.1.3. Procedure

The in-class procedure for displaying clicker questions was the same as in experiment 1, with one alteration. In the problem-oriented course, the correct answer was not offered right after voting. After projecting the class voting results, the instructor had the students discuss the question as a group and then answer a second time. After the second voting results were projected, he followed the same procedure as in the didactic course.

4.2. Results and discussion

Contrary to hypothesis 1, a repeated measures ANOVA in the physics course showed no significant effect of treatment conditions on the factual exam questions, $F(3, 399) = 1.6$, $p > 0.05$. The means (and standard deviations) for each condition are provided in Table 3. This result contrasts starkly with experiment 1's results, as well as the bulk of previously reported results of clicker effects. The bulk of prior work, however, was conducted in lecture-based, didactic classrooms and laboratories.

We also hypothesized that clicker questions would not boost performance on the conceptual exam questions. While this hypotheses was supported, an ANOVA (with a Greenhouse-Geisser correction) revealed a significant effect on conceptual exam performance, $F(2.5, 337.2) = 10.3$, $p < 0.01$, $\eta_p^2 = 0.07$. A series of paired t-tests, summarized in Table 4, revealed that conceptual clicker questions did not boost performance on conceptual exam questions. We propose the students were learning more from the weekly problem sets and laboratory inquiries than they were from the conceptual clicker intervention, as predicted. An unexpected finding was that factual clicker questions actually impeded performance, as did the enhanced control condition. It appears that focusing students' attention on the surface level of the information in the factual clicker condition drew students' attention from the underlying, more conceptual aspect of the material. The enhanced control condition had the same effect, as the attention flags drew attention to factual content rather than conceptual meaning.

Our final analysis, performed to test hypothesis 3 that student variables will affect learning outcomes, supports that interpretations. Specifically, the analysis revealed a significant 3-way interaction between clicker condition, test type and prior knowledge, $F(2.1, 271.6) = 4.7$, $p < 0.01$. As Fig. 3 shows, students who reported low prior knowledge of the course material (i.e., true beginners) performed much worse on the conceptual exam questions when in the factual clicker or enhanced control conditions. It appears, then, that the effect of those conditions on conceptual exam performance stems largely from the low prior knowledge group. For these students, focusing their attention on the surface content of course material may divert them from the deeper, more conceptual aspects of the content. Students who had some base knowledge of the content had sufficient understanding to better move beyond rote memorization. Analyses of the other student variables showed no significant interactions between the experimental conditions and students' metacognitive self-regulation, active learning, shallow or deep learning motives and strategies, or GPA, $p > 0.05$ in all cases.

5. Conclusions

Our hypotheses were partially supported by the findings. We predicted that factual and conceptual clicker questions would boost performance on factual exam questions in both courses. This hypothesis was supported in the lecture-based course in experiment 1. Neither type of clicker quested improved factual exam question performance in the problem-oriented course in experiment 2, however. We also predicted that conceptual clicker questions would boost conceptual exam performance in the didactic course. This hypothesis was not supported. We predicted that clickers would be ineffective

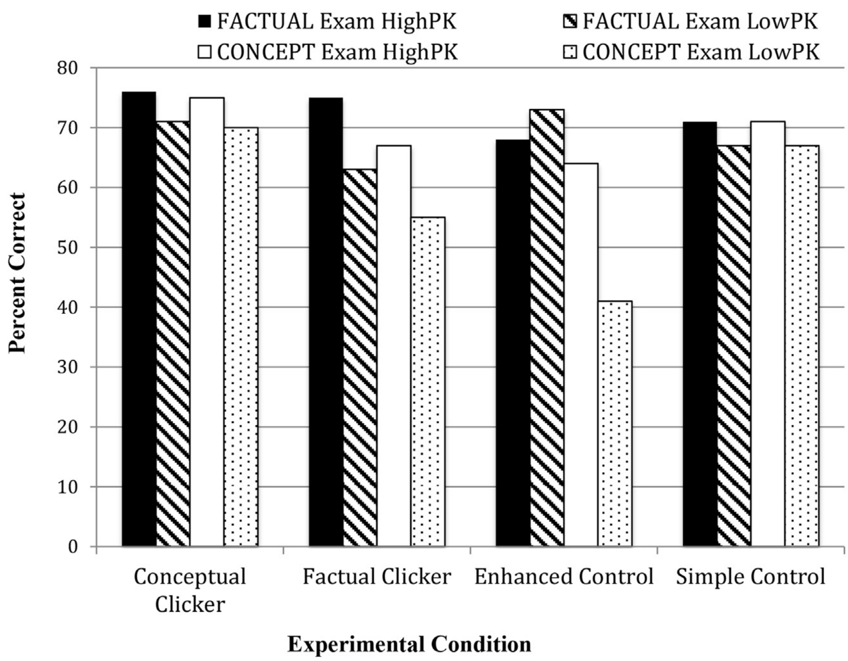


Fig. 3. Mean performance on the factual and conceptual exam questions by students in experiment 2, with high versus low prior knowledge in each of the four experimental conditions.

in the problem-oriented course, and this result was evident in the analysis. Whereas conceptual clicker questions did not improve conceptual exam performance, factual clicker questions and the enhanced control condition actually reduced performance on that measure. Finally, we predicted that subject variables would mediate clicker effects and they did. Students who were not deep learners in the didactic course performed at a level comparable to the deep learners on both types of exam questions when exposed to the clicker conditions, and worse than deep learners when in the simple control condition. Thus, clicker questions boosted students lacking deep motivation and strategies to the level of their deep motive and deep strategy peers. Also, students in the problem-oriented course with little or no prior knowledge of the material suffered more from the negative effects of the factual clicker questions and enhanced control condition on the conceptual exam questions. Prior knowledge may have served as a protective factor, preventing students from becoming diverted by the surface features of the information at the expense of developing more conceptual understanding.

We began this investigation with the goal of better understanding the conditions in which clicker use may augment learning in college courses. While many published reports indicate the technology can substantially benefit learners, we found that clicker effects are somewhat more complicated than previously reported. The technology's use appears to interact strongly with overall pedagogy, resulting in different outcomes for students enrolled in large, lecture-based courses, than for those in smaller, problem-oriented courses. Specifically, the data suggest clicker questions are effective for promoting factual learning in lecture-based, didactic courses but are not particularly effective for imparting conceptual understanding in such courses. We propose the number of conceptual clicker questions offered during class created insufficient activity to support this kind of learning in our study, and that improvement to conceptual understanding in lecture-based courses is unlikely to be seen through the adoption of clickers alone. Nor do conceptual clicker questions appear to add measurable value to courses that emphasize conceptual understanding through other activities. This finding supports Mazur's (2009) insistence that the learning enhancements he reports in his courses are not due to mere technological innovation, but to the additive effects of multiple strategies for involving students in a meaningful way with the course material. As he has stated, “it is not the technology but the pedagogy that matters (p. 51).”

The fact that the results of the enhanced control condition mirrored those of the factual clicker condition in both experiments is telling. It suggests that the reason behind clicker effects, or at least a part of the reason, is that they may alert students to important information and thus lead students to focus more on that information, either in class or during study. A study by Shapiro (2009) examined this question specifically and found that clicker effects could not be explained through an attention grabbing mechanism alone. Thus, while attention grabbing may be a factor in the clicker effects observed here, it may not account for the entire phenomenon.

An important finding that we did not anticipate before the study began was the negative effect of factual clicker questions and the enhanced control condition on conceptual understanding in the problem-oriented course. This is a novel finding in the literature, as we are aware of no prior study that documents a decrease in conceptual understanding as a result of factual in-class clicker questions. We propose the emphasis on the surface level of the material in these conditions diverted students' attention from the deeper, more conceptual level of the material. This explanation is supported by the reduced magnitude of the effect among students with greater prior knowledge. For those students, prior experience with the material may have protected students by allowing them to better recognize the important, underlying meaning of the material, and how it relates to other concepts in the course. This suggestion, of course, requires validation through future investigation. The finding, however, does suggest practitioners may be well served to consider the use of nature and use of clicker questions in the context of their goals and other facets of their pedagogy.

There are important limitations to the present study that should be noted. Most importantly, the pedagogical style we examined in each course was confounded with domain of study. Thus, it is possible that the differential effects of clickers in each course may not be related to pedagogy, but to differences between biology and physics learning. While there is no empirical evidence to support systematic domain differences in clicker-assisted learning, the possibility cannot be discarded here. Also, because the courses were taught in different domains by different instructors, with different sets of students and materials, it was not possible to make direct comparisons of learning outcomes between courses. A future study that allows such direct comparisons would be an important contribution to the findings offered here. Specifically, a profitable next step will be to test whether these differences would still be apparent if problem-solving and didactic courses were compared within a discipline.

One reviewer raised the possibility that the reuse of questions between semesters may have threatened the validity of the study if students were able to obtain copies of past exams after the first semester. In the problem-oriented course, the instructor retained the target exam questions between semesters but changed the numbers, so a memorized answer from a past test would not help. Nonetheless, we calculated mean percent correct for the experimental items in each semester. In semesters 1–4, the mean percent correct on items in the problem-oriented course was almost identical, at 68, 68, 69, and 70, respectively. In the didactic course the means were 64, 64, 61, and 60, respectively, which reflects a downward trend across successive semesters. Thus it does not appear that dishonesty among students was a significant factor in the experimental outcomes.

The present study replicates many prior reports of clicker use, which demonstrated that the technology is effective for supporting factual knowledge retention in lecture-based classrooms, but also demonstrated that the effect does not always generalize to courses employing active learning strategies. It also showed that clicker use alone may not promote conceptual understanding in either lecture-based or problem-oriented courses. It is important to note that we do not suggest instructors using active learning strategies eliminate clickers from their toolbox. Clickers can still be useful as a mechanism for soliciting student input during problem-solving exercises, group discussions, or other activities. Since there were no deleterious effects of offering conceptual clicker questions in either course, the technology may be useful for promoting group discussion, attendance, or attention. There may also be ways of presenting questions, or different types of questions unexplored in this study that will prove beneficial in the future.

Our exploration of individual differences as a contributor to learning outcomes proved fruitful, as we were able to show a relationship between study strategies and learning outcomes in the didactic course. We found that the clicker questions boosted students who did not report using deep strategies to the level of the deep strategy learners on exam questions. Many students are poor at regulating their learning, having limited ability to assess what they know and what is helpful to them. We report evidence here that clicker questions may assist them in classroom environments, in which the instructor does not promote other deep learning strategies. Prior work has indicated that students who fail to use strategies directed at deep understanding often do not know how, and interventions designed to teach those skills can be very effective. This interpretation is strengthened by the nonsignificant interaction between shallow processes and the clicker conditions. That finding suggests that students who scored low on the deep processing measure were simply not engaged in deep processes that would have helped them learn. Given this result, it may be useful to explore the interaction between metacognitive skill and clicker use. Perhaps increasing students' knowledge about their own learning behavior will affect clicker-assisted learning outcomes.

Our results may be viewed as a demonstration of the constructivist framework, and of the power of active learning, in particular. We suggest that, while clickers are useful in motivating students to come to class, increasing enjoyment of the class, and enhancing rote learning in didactic courses, instructors interested in imparting deeper understanding must be mindful of their overall pedagogy. Incorporating activities that involve students in active inquiry and problem-solving may be much more helpful than simply offering clicker questions in class, even when the clicker questions are conceptual in nature. Indeed, Mazur (2009) has cautioned against using clickers in isolation without implementing other in-class and out-of-class strategies designed to impart deep understanding. It was our intention to directly test the effects of the technology when embedded in classrooms using differing pedagogies. The present study supports Mazur's claim by demonstrating that clicker use is valuable primarily in a context in which existing pedagogy does not support deep learning. By isolating clicker-targeted material from other course material, we were able to demonstrate that clickers may not measurably augment learning for students being taught in a problem-oriented learning environment, and in some cases may detract from it.

This result also supports conclusions drawn by Levesque (2011), who found that adding clicker questions to a course that used homework assignments to provide problem-solving practice did not improve exam performance. It extends her results, however, to show that, when the in-class focus on conceptual understanding was temporarily interrupted with factual clicker questions, which focused students on surface features of the information, performance actually suffered. That a single factual question at the crucial moment in which the material was introduced in the problem-oriented course had a negative effect on conceptual understanding is an important finding. The nature, amount and timing of clicker questions may all be factors in supporting or thwarting conceptual understanding.[3]

Funding

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A100625 to the University of Massachusetts Dartmouth. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

APPENDIX A

The mean difference between the judges' factual and conceptual ratings in experiment 1 was less than 0.5, but the two physicists recruited for experiment 2 disagreed by as much as 3–4 points on individual items. Despite the operational definition supplied in the training materials, they became engaged in a very long and spirited discussion about what constituted a factual versus conceptual question in the discipline. Rater 2 had particular difficulty in adopting the operational definition for a conceptual item. Unable to persuade one another of their respective positions on multiple philosophical disagreements and, in spite of the instructions provided by the experimenter to use the provided operational definition as a guideline, they used different criteria to evaluate the materials and were not able to agree on the conceptual and factual designation of items.

To determine whether the conceptual items were inherently ambiguous, we brought in a third content expert to rate the materials on those scales. The third independent rater was also a physics professor with a doctorate, had years of experience teaching first and second year physics, and was given the same training as raters 1 and 2. The mean ratings of the factual items on the factual scale for raters 1, 2 and 3, respectively, were 4.7, 5.9, and 5.7. The mean was 5.5. The mean ratings for the conceptual items on the conceptual scale were 5.5, 3.6 and 6.1, respectively. The mean was 5.1. Rater 2 differed with rater 1 a

[3] A third course was initially included in the study. Instructor practices within that course (but outside the scope of the experiment) that came to light after the study had begun, introduced unintentionally confounding factors and error, which compromised the validity of the study. Thus, data from that course are not reported here.

mean of only 0.2 on factual items, but he differed with rater 3 by a mean of 1.3 points. He differed with raters 1 and 3 by 2.4 and 1.8 points on conceptual items, respectively. Raters 1 and 3 differed by a mean of 0.6 and 1.0 on conceptual and factual ratings, respectively. Given that all of the raters agreed on the factual items and that at least 3 of 4 physicists (three raters and the instructor) seemed to be in general agreement about what was a factual versus conceptual item, and that the items were rated as clear and of high quality, proceeded with the study using these items sets. A sample item set is provided in Fig. 1.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research, 85*, 275–314. http://dx.doi.org/10.3102/0034654314551063.

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist, 40*, 199–209. http://dx.doi.org/10.1207/s15326985ep4004_2.

Azevedo, R. (2015). Defining and measuring engagement and learning in science: Conceptual, theoretical, methodological, and analytical issues. *Educational Psychologist, 50*, 84–94. http://dx.doi.org/10.1080/00461520.2015.1004069.

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. New Directions For Teaching and Learning Series, No. 68. In L. Wilkerson, & W. Gijselaers (Eds.), *Bringing problem-based learning to higher education: Theory and practice* (pp. 3–11). San Francisco: Jossey-Bass.

Beekes, W. (2006). The 'Millionaire' method for encouraging participation. *Active Learning in Higher Education, 7*, 25–36. http://dx.doi.org/10.1177/1469787406061143.

Biggs, J. B. (1987). *The study process questionnaire (SPQ): Manual*. Hawthorn, Vic.: Australian Council for Educational Research.

Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor study process questionniare: R-SPQ-2F. *British Journal of Educational Psychology, 71*, 133–149. http://dx.doi.org/10.1348/000709901158433.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palinscar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist, 26*(3/4), 369–398. http://dx.doi.org/10.1080/00461520.1991.9653139.

Boscardin, C., & Penuel, W. (2012). Exploring benefits of audience-response systems on learning: A review of the literature. *Academic Psychiatry, 36*, 401–407. http://dx.doi.org/10.1176/appi.ap.10080110.

Brickman, P. (2006). The case of the Druid Dracula. *Journal of College Science Teaching, 36*(2), 48–53.

Choy, J. L. F., O'Grady, G., & Rotgans, J. I. (2012). Is the Study Process Questionnaire (SPQ) a good predictor of academic achievement? Examining the mediating role of achievement-related classroom behaviours. *Instructional Science, 40*, 159–172. http://dx.doi.org/10.1007/s11251-011-9171-8.

Classroom Response System ("Clickers") Bibliography. (2016). Retrieved from http://cft.vanderbilt.edu/docs/classroom-response-system-clickers-bibliography/.

Cleary, A. M. (2008). Using wireless response systems to replicate behavioral research findings in the classroom. *Teaching of Psychology, 35*(1), 42–44. http://dx.doi.org/10.1080/00986280701826642.

Crouch, C., & Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American Journal of Physics, 69*, 970–977. http://dx.doi.org/10.1119/1.1374249.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning, 20*, 81–94. http://dx.doi.org/10.1111/j.1365-2729.2004.00074.x.

English, M. C., & Kitsantas, A. (2013). Supporting student self-regulated learning in problem- and project-based learning. *Interdisciplinary Journal of Problem-based Learning, 7*(2). http://dx.doi.org/10.7771/1541-5015.1339.

Epstein, M. L., Lazarus, A. D., Calvatano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., et al. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record, 52*, 187–201.

Eysink, Tessa H. S., de Jong, Ton, Berthold, Kirsten, Kolloffel, Bas, Opfermann, Maria, & Wouters, Pieter (2009). Learner performance in multimedia learning Arrangements: An analysis across instructional approaches. *American Educational Research Journal, 46*, 1107–1149. http://dx.doi.org/10.3102/0002831209340235.

Garcia, T., & Pintrich, P. R. (1996). Assessing students' motivation and learning strategies in the classroom context: The motivated strategies for learning questionnaire. In M. Birenbaum, & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge*. Boston: Kluwer. http://dx.doi.org/10.1007/978-94-011-0657-3_12.

Gormally, C., Brickman, P., Hallar, B., & Armstrong, N. (2009). Effects of inquiry-based learning on students' science literacy skills and confidence. *International Journal for the Scholarship of Teaching and Learning, 3*(2). http://dx.doi.org/10.20429/ijsotl.2009.030216.

Greene, B. (2015). Measuring cognitive engagement with self-report scales: Reflections from over 20 years of research. *Educational Psychologist, 50*, 14–30. http://dx.doi.org/10.1080/00461520.2014.989230.

Hatch, J., Jensen, M. S., & Moore, R. (2005). 'Manna from Heaven or "clickers" from Hell: Experiences with an electronic response system. *Journal of College Science Teaching, 34*, 6–39. http://dx.doi.org/10.5480/1536-5026-33.1.55.

Herried, C. (2006). "Clicker" cases: Introducing case study teaching into large classrooms. *Journal of College Science Teaching, 36*, 43–47.

Hill, P. W., & McGraw, B. (1981). Testing the simplex assumption underlying Bloom's Taxonomy. *American Educational Research Journal, 18*, 93–101.

Jonassen, D. H., & Hung, W. (2008). All problems are not Equal: Implications for problem-based learning. *Interdisciplinary Journal of Problem-based Learning, 2*. http://dx.doi.org/10.7771/1541-5015.1080.

Kalyuga, S. (2012). Role of prior knowledge in learning processes. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2886–2888). USA: Springer. http://dx.doi.org/10.1007/978-1-4419-1428-6_214.

Kember, D., Charlesworth, M., Davies, H., McKay, J., & Stott, V. (1997). Evaluating the effectiveness of educational innovations: Using the Study Process Questionnaire to show that meaningful learning occurs. *Studies in Educational Evaluation, 23*, 141–157. http://dx.doi.org/10.1016/S0191-491X(97)00009-6.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, Cognition, 31*, 187–194. http://dx.doi.org/10.1037/0278-7393.31.2.187.

Kornell, N., & Bjork, R. A. (2009). A stability bias in human Memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General, 138*, 449–468. http://dx.doi.org/10.1037/a0017350.

Krathwohl, David R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice, 41*, 212–218. http://dx.doi.org/10.1207/s15430421tip4104.

Lambert, C. (2012). *Twilight of the lecture. Harvard magazine*. Retrieved from http://harvardmagazine.com/2012/03/twilight-of-the-lecture.

Levesque, A. (2011). Using clickers to facilitate problem-solving skills. *Cell Biology Education, 10*(4), 406–417. http://dx.doi.org/10.1187/cbe.11-03-0024.

Lin, Yi-Chun, Liu, Tzu-Chien, & Chu, Ching-Chi (2011). Implementing clickers to assist learning in science lectures: The clicker-assisted conceptual change model. *Australasian Journal of Educational Technology, 27*, 979–996. doi: 01/2011;27:979–996.

Linn, M. C., Songer, N. B., & Eylon, B. S. (1996). Shifts and convergences in science learning and instruction. In R. Calfee, & D. Berliner (Eds.), *Handbook of educational psychology*. New York: Macmillan.

Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., et al. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology, 34*(1), 51–57. http://dx.doi.org/10.1016/j.cedpsych.2008.04.002.

Mazur, E. (2009). Farewell, lecture? *Science, 323*, 50–51. http://dx.doi.org/10.1126/science.1168927.

Metcalfe, J. (1998). Cognitive optimism: Self-deception or memory-based processing heuristics? *Personality & Social Psychology Review, 2*, 100–110. http://dx.doi.org/10.1207/s15327957pspr0202_3.

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. http://dx.doi.org/10.1080/00986280701818516.

Pintrich, P. R., & Schrauben, B. (1992). Students motivational beliefs and their cognitive engagement in classroom academic tasks. In D. Schunk, & J. Meece (Eds.), *Student perceptions in the classroom: Causes and consequences*. Hillsdale, NJ: Erlbaum.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement, 53*, 801–813. http://dx.doi.org/10.1177/0013164493053003024.

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. http://dx.doi.org/10.1080/00986280701498665.

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

Rimer, S. (2009). *At M.I.T., large lectures are going the way of the blackboard. New York Times*. Retrieved from http://www.nytimes.com/2009/01/13/us/13physics.html?pagewanted=all&_r=0.

Savery, J. R. (2006). Overview of problem-based Learning: Definitions and distinctions. *Interdisciplinary Journal of Problem-based Learning, 1*(1), 9–20. http://dx.doi.org/10.7771/1541-5015.1002.

Schmidt, H. G., Loyens, S. M. M., van Gog, T., & Paas, F. (2007). Problem-based learning is compatible with human cognitive architecture: Commentary on Kirschner, Sweller and Clark (2006). *Educational Psychologist, 42*(2), 91–97. http://dx.doi.org/10.1080/00461520701263350.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460–475. http://dx.doi.org/10.1006/ceps.1994.1033.

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*(1), 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. http://dx.doi.org/10.1002/acp.2843.

Shapiro, A. M., & Gordon, L. (2013). Classroom clickers offer more than repetition: Converging evidence for the testing effect and confirmatory feedback in clicker-assisted learning. *Journal of Teaching and Learning with Technology, 2*(1), 15–30.

Shapiro, A. (1999). The relationship between prior knowledge and interactive overviews during hypermedia-aided learning. *Journal of Educational Computing Research, 20*, 143–167.

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

Stowell, J. R., & Nelson, J. M. (2007). Benefits of electronic audience response systems on student participation, learning, and emotion. *Teaching of Psychology, 34*(4), 253–258. http://dx.doi.org/10.1080/00986280701700391.

Strobel, J., & van Barneveld, A. (2009). When is PBL more Effective? A meta-synthesis of meta-analyses comparing PBL to conventional classrooms. *Interdisciplinary Journal of Problem-based Learning, 3*. retrieved from http://dx.doi.org/10.7771/1541-5015.1046.

Trees, A., & Jackson, M. (2007). The learning environment in clicker classrooms: Student processes of learning and involvement in large university-level courses using student response systems. *Learning, Media and Technology, 32*, 21–40. http://dx.doi.org/10.1080/17439880601141179.

Watkins, D. (2001). Correlates of approaches to learning: A cross-cultural meta-analysis. In R. Sternberg, & L. Zhang (Eds.), *Perspective on thinking, learning, and cognitive styles* (pp. 165–195). New Jersey: Erlbaum.

Wilder, S. (2015). Impact of problem-based learning on academic achievement in high school: A systematic review. *Educational Review, 67*, 414–435. http://dx.doi.org/10.1080/00131911.2014.974511.

Williams, J., & Lombrozo, T. (2013). Explanation and prior knowledge interact to guide learning. *Cognitive Psychology, 66*, 55–84. http://dx.doi.org/10.1016/j.cogpsych.2012.09.002.

Wirkala, C., & Kuhn, D. (2011). Problem-based learning in K–12 education. *Is It Effective and How Does It Achieve Its Effects? American Educational Research Journal, 48*(5), 1157–1186. http://dx.doi.org/10.3102/0002831211419491.