Contemporary Educational Psychology 34 (2009) 51–57

Contents lists available at ScienceDirect

Contemporary Educational Psychology

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

Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes ☆

Richard E. Mayer *, Andrew Stull, Krista DeLeeuw, Kevin Almeroth, Bruce Bimber, Dorothy Chun, Monica Bulger, Julie Campbell, Allan Knight, Hangjin Zhang

Department of Psychology, University of California, Santa Barbara, CA 93106-9660, USA

A R T I C L E I N F O

Article history:
Available online 26 August 2008

Keywords:
Educational technology
Computer-supported instruction
Post-secondary education

A B S T R A C T

What can be done to promote student–instructor interaction in a large lecture class? One approach is to use a personal response system (or "clickers") in which students press a button on a hand-held remote control device corresponding to their answer to a multiple choice question projected on a screen, then see the class distribution of answers on a screen, and discuss the thinking that leads to the correct answer. Students scored significantly higher on the course exams in a college-level educational psychology class when they used clickers to answer 2 to 4 questions per lecture (clicker group), as compared to an identical class with in-class questions presented without clickers (no-clicker group, $d = 0.38$) or with no in-class questions (control group, $d = 0.40$). The clicker treatment produced a gain of approximately 1/3 of a grade point over the no-clicker and control groups, which did not differ significantly from each other. Results are consistent with the generative theory of learning, which predicts students in the clicker group are more cognitively engaged during learning.

1. How to encourage student participation in large lecture classes

Consider the following scenario. At a large public university, 120 students are seated in a lecture hall as a professor delivers a 75-min lecture. Occasionally, the professor pauses and asks for questions or comments, but only one or two students raise their hands. The interactions between the professor and students are brief and most of the other students seem to engage in non-class related behaviors such as talking amongst themselves until the instructor returns to lecturing. This scenario is repeated for each of the 20 class meetings of the course throughout the 10-week quarter.

What is wrong with this scenario? Today, many college courses are taught in large lecture halls that hold hundreds of students. Instructors of large lecture courses may be concerned that this learning environment can lead students to feel they are passive recipients of the instructor's lecture rather than active participants in a student–instructor interaction. If students do not feel they are involved in the learning situation, they are less likely to work hard to make sense of the presented material and therefore less likely to perform as well as they could on assessments measuring their learning. What is needed is an instructional method that will engage learners in large lecture courses, allowing them to experience some degree of interaction with the instructor. Thus, part of the instructor's task is to create a sense of student–instructor interaction in a large lecture class.

2. Using questioning methods to foster learning

One way to create a feeling of student–instructor interaction in one-on-one or small group teaching situations is through a questioning method of instruction: the instructor occasionally asks a question, the student answers, and the instructor and student explain the rationale for the correct answer. We are particularly interested in using questioning methods to promote generative learning—active processing in the learner during learning such as attending to relevant material, mentally organizing the selected material, and integrating the organized material with prior knowledge. Several research literatures are relevant to the questioning method of instruction—research on adjunct question effects, research on testing effects, and research on self-explanation effects.

First, research on adding adjunct questions to printed text has shown that students perform better on a final test if they must answer adjunct questions during reading a text than if they read the text without adjunct questions (Anderson & Biddle, 1975; Andre, 1979; Andre & Theiman, 1988; Duchastel & Nungester, 1984; Mayer, 1975; McConkie, Rayner, & Wilson, 1973; Rickards &

☆ This project was supported by a grant from the Andrew W. Mellon Foundation. We wish to thank George Michaels and J.O. Davis from the UCSB Office of Instructional Development for their technical support and advice.

* Corresponding author. Fax: +1 805 893 4303.
E-mail address: mayer@psych.ucsb.edu (R.E. Mayer).

0361-476X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.cedpsych.2008.04.002

DiVesta, 1974; Rothkopf, 1966; Rothkopf & Bisbicos, 1967). In particular, classic research on adjunct questions in text has implications for the placement and type questions (Hamaker, 1986). Concerning placement of adjunct questions, students tend to perform better on tests of incidental learning (i.e., test items covering content that is different from the content in the adjunct questions) when adjunct questions are placed after rather than before the lesson (Rothkopf, 1966; Rothkopf & Bisbicos, 1967). Concerning type of adjunct questions, students tend to perform better on tests of incidental learning when the adjunct questions are conceptual questions rather than factual or verbatim questions (Mayer, 1975; Sagerman & Mayer, 1987). In order to maximize the effectiveness of questioning in the present study we placed questions after rather than before the relevant portion of the lecture, and we used conceptual questions rather than factual questions. For example, in the present study, we used conceptual questions in a multiple-choice format in which we asked students to select a prediction based on a theory rather than to simply to select the correct statement of a theory, or to select an item that describes an example of a term rather than simply to select the correct definition of the term.

More recently, research on elaborative interrogation has shown that students perform better on a final test if they must answer questions about the text material they are reading (Dornisch & Sperling, 2006; Ozgungor & Guthrie, 2004; van den Broek, Tzeng, Risden, Trabasso, & Basche, 2001; Wood, Pressley, & Winne, 1990). Some studies use questions that require a shallow level of inference and a final test that focuses mainly on recall of facts, which is not directly relevant to the present study; in contrast, other studies use questions that require a deep level of inference and a final test that goes beyond recall of facts (Dornisch & Sperling, 2006; Ozgungor & Guthrie, 2004), which is consistent with the present research. Teaching students how to ask questions during learning is another effective way to promote generative learning (King, 1992; Rosenshine, Meister, & Chapman, 1996; Wiser & Graesser, 2007), although teaching of learning strategies was not our focus in this study.

Second, research on the testing effect has shown that students perform better on a final test if they take a practice test (without feedback) on a lesson they have received rather than restudy the lesson (Foos & Fisher, 1988; Roediger & Karpicke, 2006). There is consistent support for the testing effect across many experiments dating back to the early 1900s, especially when the final test was a delayed retention test (Roediger & Karpicke, 2006). Third, research on the self-explanation effect has shown that students perform better on a final test when they are encouraged to explain aloud to themselves as they read a textbook lesson rather than simply read the lesson without engaging in self-explanation (Roy & Chi, 2005).

The rationale for each of these manipulations is that it fosters generative learning, leading to superior test performance. In short, it appears that generative methods of instruction—such as adjunct questions, practice testing, and self-explanation—can be effective, particularly for learning of verbal material. Although all of these literatures encourage the present study, none of them focuses specifically on questioning methods in large lecture courses. In the present study, we examine whether questioning can be used successfully to foster generative learning in a large lecture class.

3. How to implement questioning methods in large lecture classes

An important challenge is to incorporate the benefits of a questioning method of instruction in a large lecture class. One proposed solution to this problem is to take advantage of newly emerging educational technologies that purport to allow for learner interactivity in large lecture courses, and thereby foster better learning. In particular, proponents have proposed using a personal response system (or "clickers") in which students press a button on a hand-held remote control device corresponding to their answer to a multiple choice question that is being projected on a screen, see the correct answer along with the class distribution of answers, and hear a description of the thinking that leads to the correct answer (Duncan, 2005). In the present studies, a clicker-based system was used to present 2 to 4 multiple-choice questions during each lecture, ask students to vote using their hand-held clickers, and then in a matter of seconds show a graph indicating the correct answer along with the percentage of students who voted for each answer alternative. Then, the instructor called on one student to explain the correct answer and finally the instructor described his thought process leading to the correct answer. In short, the instructional technology of clickers was used to implement the instructional method of questioning.

Although personal response systems seem promising, limited research has been conducted on their effectiveness in implementing a questioning method in college courses. Much of the research on clickers in the classroom has focused less on learning outcomes and more on self-reports of how helpful the students found the remote controls or how much they enjoyed using them (Beekes, 2006; Draper & Brown, 2004; Duncan, 2005; Hatch, Jensen, & Moore, 2005; Latessa & Mouw, 2005; Wit, 2003; Zahorik, 1996). Duncan (2005, p. 22) has claimed that "proper clicker use can lead to higher grades," but offers no published peer-reviewed evidence to support the claim. Informal studies of the instructional effectiveness of clickers are difficult to interpret because they lack control groups (Duncan, 2005). Recent surveys of students' experiences in learning with clickers (Trees & Jackson, 2007) and teacher's experiences in teaching with clickers (Penuel, Boscardin, Masyn, & Crawford, 2007) provide interesting information concerning the self-reported benefits of clickers, but only experimental comparisons allow for causal conclusions concerning effects on learning outcomes (Phye, Robinson, & Levin, 2005). In spite of strong claims and high hopes expressed in the literature, our search for peer-reviewed data to use in a meta-analysis on learning effect sizes yielded no results. Overall, we were unable to identify any peer-reviewed published articles comparing a clicker group to a control group on a learning test.

The present 3-year study seeks to produce a methodologically sound and ecologically valid test of the pedagogic value of an instructional method implemented by using clickers. In particular, we investigated the exam performance of students who took a college course in educational psychology, comparing those who experienced a clicker-supported questioning method (clicker group) to those who experienced in-class questioning implemented without clickers (no-clicker group) and others who experienced no in-class questioning or clickers (control group).

4. A generative model for clicker-based instructional methods

How does asking questions produce student learning? According to the generative theory of learning, students learn better when they engage in active cognitive processing during learning (Mayer & Wittrock, 2006; Wittrock, 1990). In generative theory, the learner's behavioral activity during learning does not cause learning but rather the learner's cognitive activity during learning causes learning. Mayer (2001, 2008) has identified three cognitive processes involved in generative learning: selecting the relevant material from the incoming lesson, organizing the selected material into a coherent representation in working memory, and integrating the representation with existing knowledge from long-term memory. For example, in a lecture on educational psychology, students must focus on the relevant aspects of what the instructor is saying such as the key points in a description of a research study; students must mentally organize the material into a coherent structure such as a schema consisting of method, results, and conclusion; and must mentally connect the incoming material with prior knowledge, perhaps about a similar experiment.

According to generative theory, certain instructional methods can prime these cognitive processes during learning (Mayer, 2008; Mayer & Wittrock, 2006). In this study we focus on the instructional method of questioning as a technique intended to prime active cognitive processing in learners. In particular, in the questioning treatments we present 2 to 4 multiple-choice questions per lecture based on the lecture content, ask all students to respond, show how many students selected each alternative, and discuss the rationale for the correct answer. Questioning can be a generative method of instruction because when students answer questions during learning they are encouraged to select relevant information, mentally organize the material, and integrate it with their prior knowledge. For example, when students are asked to make predictions based on a theory, they are required to think more deeply about the theory. When asked to determine which example best matches a term, they are required to think more deeply about the definition. Experience in answering practice questions and justifying the correct answer, may encourage students to also process other course material more deeply.

According to generative theory, the outcome of active cognitive processing during learning is a meaningful learning outcome, which can be assessed through retention and transfer tests (Anderson et al., 2001; Mayer & Wittrock, 2006). Consistent with guidelines for the design of assessment of learning outcomes (Anderson et al., 2001; Pellegrino, Chudowsky, & Glaser, 2001), in the present study we evaluated learning with test items on a variety of kinds of knowledge and skills covered in the course—including items on material that is similar and dissimilar to the questions used in class.

In the present study, we attempted to create a clicker-based instructional method that emphasized the academic content—i.e., being able to answer exam-like questions. The act of trying to answer sample questions and then receiving immediate feedback may encourage active cognitive processing in three ways: (a) before answering questions, students may be more attentive to the lecture material, (b) during question answering, students may work harder to organize and integrate the material, and (c) after receiving feedback, students may develop metacognitive skills for gauging how well they understood the lecture material and for how to answer exam-like questions.

Thus, our main prediction is that the clicker treatment will lead to greater student–teacher interaction, which encourages deeper cognitive processing during learning, which in turn will be reflected in improvements in exam score in the course. In short, we expect the clicker group to produce higher exam scores than the control group. If we are successful in implementing the questioning method without computer-based technology in the no-clicker group, we also expect the no-clicker group to outperform the control on exam scores and to be equivalent to the clicker group.

5. Reasons for caution in using educational technology

The history of educational technology is replete with examples of strong claims for the pedagogic value of new technologies followed by failures of the technology to improve student learning–including claims for motion pictures in the 1920s, educational radio in the 1930s and 1940s, educational television in the 1950s, and computer-based programmed instruction in the 1960s (Cuban, 1986). The search for an appropriate educational technology can become a misleading and potentially unproductive adventure, unless we consider the fundamental distinction between instructional media and instructional method (Clark, 2001; Mayer, 2001). The instructional medium refers to the physical devices used to deliver instruction, such as using printed words and illustrations in books, handwritten words and diagrams on blackboards, PowerPoint slides on a screen, or computer-based animation and narration accessed over the internet. The instructional method refers to the techniques used to foster cognitive processing in the learner, such as providing worked-out examples, asking students to explain the material to themselves, or drill and practice. Researchers in the field of educational technology have reached consensus that media do not cause learning, but rather methods cause learning. Therefore, in this project, we focus on an instructional method that is afforded by an instructional technology rather than on the technology itself. In particular, we seek an instructional method that encourages student–instructor interaction for all students in a large lecture class. In a recent analysis of how various instructional technologies enable various types of instructional methods for promoting various cognitive processing during learning, Mayer et al. (2006) identified personal response systems as a technology that could enable instructional methods that promote student–instructor interaction during learning.

Similarly, Mayer (2001) has identified two approaches to the use of educational technology: a technology-centered approach and a learner-centered approach. In a technology-centered approach, instructional designers focus on a new technology and ask, "How can we apply this technology in education?" In short, learners and instructors are required to adapt to a new technology. In a learner-centered approach, instructional designers focus on the learner, and ask, "How can the learner's cognitive processing be aided by a new technology?" In short, the technology is adapted to become a tool that expands human cognition. In this study, we take a learner-centered approach by seeking to foster appropriate cognitive processing in students attending large lecture classes.

6. Method

6.1. Participants and design

The participants were 139 college students who completed Educational Psychology (Psychology 124) during the 2004–2005 academic year at the University of California, Santa Barbara (UCSB) in a large lecture course that did not involve technology or group questioning (control group), 111 college students who completed the same course during the 2005–2006 academic year in a large lecture course that involved technology in the form of a personal response system (clicker group), and 135 college students who completed the same course during the 2006–2007 academic year in a large lecture course that involved group questioning without clickers. All students were Psychology majors and all took both a pre-questionnaire and post-questionnaire administered in class. As shown in Table 1, 99% of the students

Table 1
Comparison of three groups on demographic characteristics

Characteristic	Clicker group	No-clicker group	Control group
Mean SAT score (and *SD*)	1145 (115)	1194 (122)	1159 (121)
Proportion of juniors and seniors	.99	.98	.99
Proportion of women	.80	.73	.73

were juniors or seniors, 75% of the students were women, and the students averaged 1167 on the SAT.

6.2. Materials and apparatus

The materials consisted of two to four PowerPoint slides for each of 18 lectures, with each slide containing a multiple choice question that covered a portion of the lecture content. The verbatim "clicker" questions were not on the exam. For example, a clicker question from an early lecture on principles of learning was:

Thorndike asked a group of students who had learned Latin and a group of students who had not taken Latin to learn a new subject such as bookkeeping. According to Thorndike's theory of transfer by identical elements, which group should learn the new subject better?

(a) Students who knew Latin will learn better because Latin fosters proper habits of mind.
(b) Students who had not taken Latin will learn better because the components in Latin conflict with the components in bookkeeping.
(c) Both will learn the same.
(d) The theory of transfer by identical elements does not make a prediction.

The materials also consisted of a midterm exam containing 45 multiple-choice questions covering the first half of the course and a final exam containing 45 multiple-choice questions covering the second half of the course. Thirty of the exam questions covered similar content to the clicker questions, although the verbatim clicker questions were not used on the exams. An example of a similar question is:

Thorndike asked a group of students who had learned Latin and a group of students who had not taken Latin to learn a new subject such as bookkeeping. According to doctrine of formal discipline, which group should learn the new subject better?

(a) Students who knew Latin will learn better because Latin fosters proper habits of mind.
(b) Students who had not taken Latin will learn better because the components in Latin conflict with the components in bookkeeping.
(c) Both will learn the same.
(d) The doctrine of formal discipline does not make a prediction.

Sixty of the exam questions covered dissimilar content to the clicker questions, although the material was covered in the course. An example of a dissimilar problem is:

Jenny is given a word problem in math class, and she immediately classifies the problem as a right triangle problem. Which type of knowledge is Jenny using to classify this problem as a right triangle problem?

(a) procedural
(b) situational
(c) strategic
(d) schematic

This problem is dissimilar because none of the clicker questions asked about the types of knowledge used in solving mathematics problems.

The materials also consisted of a pre-questionnaire that solicited basic demographic information and a post-questionnaire that solicited self-reported course-related activities.

The apparatus consisted of the TurningPoint (2005) personal response system, which included 150 radio frequency (RF) response transmitters, a radio frequency (RF) receiver, and a Sony Vaio laptop computer running TurningPoint and PowerPoint software. The receiver was connected to the computer through the USB port.

6.3. Procedure

We used a quasi-experimental design in which we compared the combined midterm and final exam scores of students who took an educational psychology course in 2005 (control group) with those taking the same course in 2006 (clicker group) and those taking the same course in 2007 (no-clicker group). The instructor, student eligibility requirements, lecture content, reading assignments, and exam questions were identical in the three classes. In all classes, the pre-questionnaire was administered during the first week of class and post-questionnaire was administered during the last week of class. The pre-questionnaire and post-questionnaire were anonymous and were linked to student performance records by code numbers rather than names. In all classes, the midterm exam covered the first five weeks of the course and was given during the sixth week of class and the final exam covered the final five weeks of the course and was given during exam week which occurred after 10 weeks of class. In all classes students were asked to sign a consent form during the first week of class.

The major difference[1] among the three classes concerned the way the instructor interacted with the class. The clicker group received approximately 5 to 10 min per lecture devoted to answering and discussing 2 to 4 questions multiple-choice questions presented by the instructor. In the clicker treatment, students pressed a button on a hand-held remote control device corresponding to their answer to a multiple choice question projected on a screen, saw the class distribution of answers with the correct answer indicated, and heard the thinking that leads to the correct answer.

In the clicker group, during the first week of class, each student was given a transmitter for use throughout the quarter and the transmitter's identification number was registered to the student in the instructor's computer-based database. In each lecture, after a section of the lecture, the instructor presented one or more sample multiple choice questions on the screen using PowerPoint presentation software in conjunction with TurningPoint. The instructor introduced the question by saying something like, "OK, now let's see how well you understood the material we just covered. Please take out your clickers and press a, b, c, or d." When all students had pressed a button on their response transmitters (which generally took 20 to 30 s), the instructor displayed a graph showing the correct answer and the percentage of students who selected each answer. Then, a short discussion ensued concerning the rationale for the correct answer, which the instructor summarized. The instructor initiated the discussion by asking a question such as, "Why is c the correct answer?" Approximately, 2 min of class time was used for each question. The TurningPoint software recorded each student's response and allocated 2 points for each correct answer or 1 point for an incorrect answer. Students could earn up to 40 points in course credit for answering the "clicker questions" in class, and could check their points on a class website. During the last week of class, students handed in their clickers.

In the no-clicker group, the instructor passed out a sheet containing 2 to 4 multiple-choice questions at the end of the lecture (or a section of the lecture), asked students to indicate their answers on a form at the bottom of the sheet, collected the forms, then asked for a raise of hands for each alternative on the first

[1] The clicker treatment also included the Moodle Course Management System, which was used to post the course syllabus; allow on-line submission of 4 course assignments (i.e., with each assignment requiring the student to write an exam question); and allow students to examine their current point count for clicker questions, exams, and assignments.

question, gave the correct answer, called on a student to explain the correct answer, and explained his thinking in arriving at the correct answer. After all questions had been covered, students were asked to score their sheet (with two points for a correct answer and one point for an incorrect answer) and the instructor collected the sheets. As in the clicker group, students could earn up to 40 points for answering questions.

In the control group, the instructor simply asked the class if they had any questions at various points in the lecture, but did not present any multiple-choice questions or give students up to 40 points for answering in-class questions.

7. Results and discussion

7.1. Were the groups equivalent on basic student characteristics?

A preliminary issue concerns whether the clicker, no-clicker, and control groups were equivalent on basic learner characteristics. Table 1 shows the mean combined SAT scores (and standard deviations), the proportion of juniors and seniors, and the proportion of women in each of the three groups. The mean combined SAT scores differed significantly among the three groups, based on an analysis of variance, $F(2,267) = 3.43$, $p = .034$. Based on a Newman–Kuels' test with alpha at .05, the no-clicker group had a significantly higher score than the clicker group, and no other differences were significant. Importantly, the group that we predicted would show the greatest learning (i.e., clicker group) did not have significantly higher SAT scores than the other groups.

The proportion of juniors and seniors did not differ significantly among the groups based on a chi square test, $\chi^2(2, N = 337) = .99$, $p = .607$. The proportion of women did not differ significantly among the three groups based on a chi square test, $\chi^2(4, N = 355) = 4.39$, $p = .36$.

7.2. Is the exam score reliable?

The main dependent measure in this study is the students' scores on the exams. Based on Cronbach's alpha, exam score reached an acceptable level of internal reliability, $\alpha = .80$. The internal reliability for similar items was $\alpha = .53$ and the internal reliability for dissimilar items was $\alpha = .63$. SAT scores correlated significantly with exam scores, $r = .28$, $p < .01$.

7.3. Does the clicker treatment improve academic performance?

The main prediction concerns the effects of the clicker treatment on academic performance. For each student, we tallied the number of correctly answered questions on the midterm exam (out of 45 questions) and the number of correctly answered questions on the final exam (out of 45 questions), yielding a total score (out of 90 possible). As shown in the first line of Table 2, the clicker group ($M = 75.1$, $SD = 6.9$) correctly answered more exam questions than did the control group ($M = 72.3$, $SD = 7.7$, $d = .38$) or the no-clicker group ($M = 72.2$, $SD = 7.6$, $d = .40$). These means correspond to 83.4% correct or a grade of B for the clicker group, 80.3% correct or a grade of B− for the control group, and 80.2% correct or a grade of B− for the no-clicker group.[2] These scores indicate a 1/3 grade point improvement for the clicker group over the other two groups (i.e., from B− to B). An analysis of variance conducted on the exam score data revealed that the groups differed significantly from one another, $F(2,382) = 5.76$, $MSE = 55.39$, $p = .003$. Supplemental pairwise comparisons using Newman–Keuls' tests, with alpha at $p < .05$, revealed that the clicker group outperformed the control and no-clicker groups, which did not differ from each other. These findings support the main prediction the clicker-supported questioning method would improve academic achievement.

Table 2
Mean score on exam questions for three groups based on all data, similar items, and dissimilar items

Partition (and total possible)	Clicker group		No-clicker group			Control group		
	M	*SD*	*M*	*SD*	*d*	*M*	*SD*	*d*
All data (90)	75.1	6.9	72.3	7.5	.39*	72.2	7.6	.40*
Similar items (30)	24.9	2.4	24.4	3.0	.18	24.2	2.9	.26
Dissimilar items (60)	50.2	5.0	47.9	5.3	.44*	48.2	5.5	.38*

Note. Asterisk (*) indicates significant difference ($p < .05$) between no-clicker and clicker groups or between clicker and control groups.

A secondary issue concerns whether questioning can be successfully implemented without using computer-based technology. In the present study, students who received questioning using a paper-based system (no-clicker group) performed at the same level as students who received no questioning (control group), indicating that the paper-based questioning treatment had no discernable effect on student learning outcomes. Why did the instructional method of questioning work when it was implemented using clickers but not work when it was implemented using paper? One possible explanation is that the logistics of using a paper-based implementation of questioning resulted in more disruption to class time than did the logistics of using a clicker-based implementation of questioning. The clicker-based implementation worked flawlessly, so the instructor could transition easily to the clicker questions—sometimes at several different points in the lecture—and easily transition back to the lecture. The students' responses were recorded automatically so there was no need to spend time collecting responses manually. In contrast, the paper-based implementation required passing out question sheets to each student in the class, collecting answer forms from each student, and collecting the question sheets from each student—all of which took considerable time and was disruptive to the flow of the lecture. All questions had to be administered at the same time, usually at the end of the lecture, which involved a long period of "quiet time" and may have created too much of a separation from the lecture. To gauge student responses, students were asked to raise their hands for each alternative (rather than view a graph), and some students did not raise their hands. In addition, in the no-clicker group students did not have immediate access to their "clicker points" on the course website but did in the clicker group.

Overall, the paper-based technology created logistical problems that may have detracted from the effectiveness of the instructional method, because it was highly disruptive to the flow of the lecture. In contrast, the clicker-based technology afforded the opportunity to use the instructional method of questioning in a way that was seamless with the lecture. Thus, although we maintain that it was the instructional method—questioning—that caused an improvement in academic achievement, it appears that the clicker-based technology was more capable of implementing the method than was the paper-based technology.

7.4. Does the clicker treatment improve academic performance on non-clicker related questions?

A possible criticism of this study may be that the clicker group was exposed to questions that were similar to those on the exam, whereas the control group was not. Importantly, as shown in the second section of Table 2, the pattern of improvement favoring the

[2] This analysis is based on the idea that from 80% up to 83% correct is a B− (or 2.7 grade points on a 4-point scale), from 83% up to 87% correct is a B (or 3.0 on a 4-point scale), and from 87% up to 90% correct is B+ (or 3.3 on a 4-point scale).

clicker group is strong for exam questions that involved dissimilar content from the in-class clicker questions, $F(2,381) = 6.47$, $MSE = 180.61$, $p = .002$. For these dissimilar items, a Newman–Kuels' test (with $p < .05$) indicated that the clicker group outscored the control group ($d = .38$) and the no-clicker group ($d = .44$), which did not differ from each other. Although the same pattern can be seen for the similar items, the differences among the groups were not statistically significant, $F(2,381) = 2.50$, $MSE = 19.60$, $p = .08$. Thus, the advantage of the clicker treatment is not restricted to the content addressed in the clicker questions, and the effects of the clicker treatment cannot be attributed solely to directing the learner's attention to specific course content.

Overall, these results show that an instructional method based on questioning and supported by clicker technology can improve student academic performance in a large lecture class. The effect of the clicker treatment on academic performance is significant both in a statistical sense—i.e., the difference was statistically significant at the .05 level—and in a practical sense—i.e., the difference produced at effect size of .38 and raised grades by approximately 1/3 grade point. In short, this project demonstrates that it is possible to successfully implement instructional methods aimed at fostering generative learning in a large lecture course.

8. Conclusion

8.1. Practical implications

Concerning educational significance, this study contributes to the fledgling literature on whether personal response systems can be used to improve students' academic performance in an authentic classroom environment. In a quasi-experimental design, we found evidence that a personal response system can be used in ways that promote academic performance in large lecture classes at the college level. In particular, the personal response system was used to stimulate student–instructor interaction concerning how to answer sample test items. Thus, our results are consistent with previous research on the benefits of using questions in text. If the goal is to help students learn in large college lecture classes, there is reason to consider using a personal response system to foster student–instructor interaction during class. Interestingly, the clicker group outperformed both the control group and the no-clicker group, suggesting that the implementation of the questioning method was less intrusive with clicker technology.

8.2. Theoretical implications

Concerning scientific significance, these results are consistent with the generative theory of learning, which proposes that students learn better when they engage in appropriate cognitive processing during learning. Our hypothesis is that the act of trying to answer sample questions and getting immediate feedback, encouraged students to engage in appropriate cognitive processing including: (a) paying more attention to the lecture in anticipation of having to answer questions, (b) mentally organizing and integrating learned knowledge in order to answer questions, and (c) developing metacognitive skills for gauging how well they understood the lecture material and for how to answer exam-like questions in the future. We hypothesized that the clicker-supported questioning method would encourage a sense of student–instructor interaction, which would lead students to try harder to make sense of the presented material. Although the present study provides some evidence that students in the clicker group learned better, the present study was not intended to provide data on differences in students' sense of student–instructor interaction or cognitive processing during learning. Further research is needed to investigate these important underlying mechanisms.

There are a number of alternative explanations, all based on the idea that the same method of instruction and effects on learning outcome could be accomplished without computer-based technology. For example, perhaps being exposed to questions like those on the exams could be all that is needed to help students adjust their study practices. Alternatively, being exposed to questions and required to answer them by raising their hands, could be all that is needed to motivate students to engage more deeply with the material. Finally, simply giving students points for attending the lectures—as was effectively done in the clicker class—may improve attendance, which would lead to better test scores. Aspects of these explanations were addressed by including the no-clicker group, in which students were exposed to the same questions as the clicker group, required to raise their hands, and received points for answering. The finding that the no-clicker group performed at the same level as the control group suggests that being able to seamlessly integrate the questioning method was an important distinguishing feature of the clicker treatment. Further research is needed to determine which features of the clicker treatment improved student learning, and whether the effects persist after the novelty of using clickers diminishes over time.

8.3. Methodological implications

Concerning methodological significance, the current study attempts to show how it is possible to conduct research involving educational technology by focusing on an educationally relevant instructional method afforded by the technology rather than by the technology per se. It also attempts to combine ecologically validity (by studying actual college courses) with methodological soundness (by using a quasi-experimental design that controls for as many potential confounds as possible). Although this study yielded encouraging results, it should be seen as a first step in investigating the pedagogical value of clicker-based instructional methods in large lecture classes. Like other methodologically rigorous research on the classroom use of technology, research on the effects of clicker-based instructional methods can have useful implications for evidence-based practice (Chambers, Cheung, Madden, Slavin, & Gifford, 2006).

8.4. Limitations and future directions

In this section, we consider three possible limitations of the study: researcher bias, novelty, and confounds. First, the potential for researcher bias is present in the current study because one of the authors (Mayer) was an instructor for the course. Even though every effort was made to equate the courses, the instructor may have unconsciously changed his teaching style in subtle ways that would support the main prediction. It would be impossible to hide the experimental treatment from the instructor, even if the instructor was not a member of the research team because it obvious that there is a difference among the treatments. However, a mitigating factor in the present study is that the instructor did not expect the clicker group to perform better than the no-clicker group or even to necessarily perform better than the control group. Thus, the instructor's expectations did not correspond to the actual results.

Second, the potential for novelty effects is present in the current study because the use of clicker technology is still a relatively new phenomenon. At the university where this study took place, the target course was the only one using clickers in the entire department. For all students in the study, this was the first time they had used clickers in a psychology course. It is possible that the novelty of using a new technology caused them to work harder in the course, and therefore was the cause the clicker effect in this study. We do not have data to address this alternative explanation, but future research should examine the effects of clickers longitudinally to determine whether the effects diminish with increasing clicker experience.

Third, additional potential confounds should be considered because the three treatments differed along several dimensions. Although our intention was to compare computer-based questioning, paper-based questioning, and no questioning methods in a college course, it is possible our obtained differences were caused by confounding factors, such as differences in the incentive to attend class to earn points, differences in student characteristics, and differences in time required to carry out the treatments. Concerning incentive to attend class, students in the clicker group received up to 40 course points for answering questions in class whereas students in the control group did not, so the difference in exam scores may be attributable to differences in the incentive to attend class. However, the no-clicker group had the same incentive structure as the clicker group, but performed like the control group, thereby suggesting that incentive structure per se was not the main causal agent in the study.

Concerning student characteristics, a quasi-experiment always runs the risk that prior differences exist between the groups on variables not measured, and that these differences cause differences in the outcome variable. In the present study, we had no reason to suspect that the students taking the course would differ from year to year, as all students were Psychology majors generally in their junior or senior year. Our analysis of basic demographic characteristics such as sex, year in school, and SAT score helped confirm this expectation. The only significant difference—in SAT score—did not favor the clicker group, thereby suggesting that prior differences in scholastic ability were not responsible for the clicker effect.

Concerning time differences for the treatments, it is clear that the control group had more time to cover course content and engage in classroom discussion than did the questioning groups, but the extra time apparently did not help the control group. It should be noted that the no-clicker treatment required more off-task time than the clicker treatment, because of the need to pass out and pass back sheets with questions. It is possible that this disruption detracted from the flow of the lecture and diminished the effectiveness of the questioning method. In short, the questioning method may have been useful but the computer-based technology enabled a less disruptive way of implementing it than did the paper-based technology. Future research is needed to pinpoint the locus of the questioning effect we obtained.

References

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al. (2001). *A taxonomy of learning for teaching: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Anderson, R. G., & Biddle, W. B. (1975). On asking people questions about what they are reading. *Psychology of Learning and Motivation, 9*, 90–132.

Andre, T. (1979). Does answering high level questions while reading facilitate productive learning? *Review of Educational Research, 49*, 280–318.

Andre, T., & Theiman, A. (1988). Level of adjunct questions, type of feedback, and learning concepts by reading. *Contemporary Educational Psychology, 13*, 296–307.

Beekes, W. (2006). The "millionaire" method for encouraging participation. *Active Learning in Higher Education, 7*, 25–36.

Chambers, B., Cheung, A. C. K., Madden, N. A., Slavin, R. E., & Gifford, R. (2006). Achievement effects of embedded multimedia in a successful for all reading program. *Journal of Educational Psychology, 98*, 232–237.

Clark, R. E. (Ed.). (2001). *Learning from media: Arguments, analysis, and evidence*. Greenwich, CT: Information Age Publishers.

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York: Teachers College Press.

Dornisch, M., & Sperling, R. A. (2006). Facilitating learning from technology-enhanced text: Effects of prompted elaborative interrogation. *Journal of Educational Research, 99*, 156–165.

Draper, S. W., & Brown, M. I. (2004). Increasing interactivity in lectures using an electronic voting system. *Journal of Computer Assisted Learning, 20*, 81–94.

Duchastel, P., & Nungester, R. J. (1984). Adjunct question effects with review. *Contemporary Educational Psychology, 9*, 97–103.

Duncan, D. (2005). *Clickers in the classroom: How to enhance science teaching using classroom response systems*. San Francisco: Pearson/Addison-Wesley.

Foos, P. W., & Fisher, R. P. (1988). Using tests as learning opportunities. *Journal of Educational Psychology, 80*, 179–183.

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56*, 212–242.

Hatch, J., Jensen, M., & Moore, R. (2005). Manna from heaven or clickers from hell: Experience with an electronic response system. *Journal of College Science Teaching, 34*, 36–39.

King, A. (1992). Comparison of self-questioning, summarizing, and note taking review as strategies for learning from lectures. *American Educational Research Journal, 29*, 303–323.

Latessa, R., & Mouw, D. (2005). Use of an audience response system to augment interactive learning. *Family Medicine, 37*, 12–14.

Mayer, R. E. (1975). Forward transfer of different reading strategies due to test-like events in mathematics text. *Journal of Educational Psychology, 67*, 165–169.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2008). *Learning and instruction*. New York: Pearson Merrill Prentice Hall.

Mayer, R. E., Almeroth, K., Bimber, B., Chun, D., Knight, A., & Campbell, A. (2006). Technology comes to college: Understanding the cognitive consequences of infusing technology in college classrooms. *Educational Technology, 46*, 48–53.

Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology: Second edition* (pp. 287–304). Mahwah, NJ: Erlbaum.

McConkie, G. W., Rayner, K., & Wilson, S. J. (1973). Experimental manipulation of reading strategies. *Journal of Educational Psychology, 65*, 1–8.

Ozgungor, S., & Guthrie, J. T. (2004). Interactions among elaborative interrogation, knowledge, and interest in the process of constructing knowledge from text. *Journal of Educational Psychology, 96*, 437–443.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know*. Washington, DC: National Academy Press.

Penuel, W. R., Boscardin, C. K., Masyn, K., & Crawford, V. M. (2007). Teaching with student response systems in elementary and secondary education settings: A survey study. *Educational Technology Research and Development, 55*, 315–336.

Phye, G. D., Robinson, D. H., & Levin, J. (Eds.). (2005). *Empirical methods for evaluating educational interventions*. San Diego: Elsevier Academic Press.

Rickards, J. P., & DiVesta, F. J. (1974). Type and frequency of questions in processing textual material. *Journal of Educational Psychology, 66*, 354–362.

Roediger, H. L., & Karpicke, J. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Rothkopf, E. Z. (1966). Learning from written materials: An exploration of the control of inspection of test-like events. *American Educational Research Journal, 3*, 241–249.

Rothkopf, E. Z., & Bisbicos, E. (1967). Selective facilitative effects of interspersed questions on learning from written material. *Journal of Educational Psychology, 58*, 56–61.

Roy, M., & Chi, M. T. H. (2005). The self-explanation effect in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 271–286). New York: Cambridge University Press.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of intervention studies. *Review of Educational Research, 66*, 181–221.

Sagerman, N., & Mayer, R. E. (1987). Forward transfer of different reading strategies evoked by adjunct questions in science text. *Journal of Educational Psychology, 79*, 189–191.

Trees, A. R., & Jackson, M. H. (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.

TurningPoint (2005). TurningPoint 2006 User Guide. Youngstown, OH: Turning Technologies.

Wiser, R. A., & Graesser, A. C. (2007). Question-asking in advanced learning environments. In S. M. Fiore & E. Salas (Eds.), *Toward a science of distributed learning* (pp. 209–234). Washington, DC: American Psychological Association.

Wit, E. (2003). Who wants to be...The use of a personal response system in statistics teaching. *MSOR Connections, 3*, 14–20.

van den Broek, P., Tzeng, Y., Risden, K., Trabasso, T., & Basche, P. (2001). Inferential questions: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology, 93*, 521–529.

Wittrock, M. C. (1990). Generative processes of comprehension. *Educational Psychologist, 24*, 354–376.

Wood, E., Pressley, M., & Winne, P. (1990). Elaborative interrogation effects on children's learning of factual content. *Journal of Educational Psychology, 82*, 741–748.

Zahorik, J. (1996). Elementary and secondary teachers' reports of how they make learning interesting. *Elementary School Journal, 98*, 3–13.