Article

Using Clickers in Nonmajors- and Majors-Level Biology Courses: Student Opinion, Learning, and Long-Term Retention of Course Material

Kirsten Crossgrove and Kristen L. Curran

Department of Biological Sciences, University of Wisconsin, Whitewater, WI 53190

Student response systems (clickers) are viewed positively by students and instructors in numerous studies. Evidence that clickers enhance student learning is more variable. After becoming comfortable with the technology during fall 2005–spring 2006, we compared student opinion and student achievement in two different courses taught with clickers in fall 2006. One course was an introductory biology class for nonmajors, and the other course was a 200 level genetics class for biology majors. Students in both courses had positive opinions of the clickers, although we observed some interesting differences between the two groups of students. Student performance was significantly higher on exam questions covering material taught with clickers, although the differences were more dramatic for the nonmajors biology course than the genetics course. We also compared retention of information 4 mo after the course ended, and we saw increased retention of material taught with clickers for the nonmajors course, but not for the genetics course. We discuss the implications of our results in light of differences in how the two courses were taught and differences between science majors and nonmajors.

INTRODUCTION

The use of student response systems in the classroom has been discussed widely in the literature, and was recently reviewed by Caldwell (2007). Briefly, student response systems (clickers) are small hand-held keypads that allow students to answer a multiple-choice (MC) question displayed on a projection system. A receiver on the instructor's computer collects the information, and it is displayed as a graph of the students' responses.

Students and instructors both tend to have positive opinions of clickers, but the data on whether clickers increase student learning are more variable (Judson and Sawada, 2002). Learning gains were generally not observed in older studies when clickers were used primarily for stimulus-response type learning (Bessler and Nisbet, 1971; Judson and Sawada, 2002). However, more recently, clickers have shown positive effects on student learning when used in conjunction with active-learning strategies such as peer instruction (Judson and Sawada, 2002; Duncan, 2005; Knight

and Wood, 2005; Caldwell, 2007; Freeman et al., 2007). In the life sciences, studies analyzing the effects of clicker use on student learning between different cohorts of students (between classes or semesters) report no effect in some studies (Paschal, 2002), whereas other studies report a positive effect on student learning (Knight and Wood, 2005; Freeman et al., 2007). Suchman et al. (2006) compared two sections of a microbiology class and observed higher exam scores in the section that used clickers more extensively. However, the increase in student performance was the same regardless of whether the questions were based on material covered by clickers or not. Studies comparing the effect of clickers on student learning in the same cohort of students show increases in some cases (Preszler et al., 2007), but not in other cases (Bunce et al., 2006). Several confounding factors make comparison of these studies difficult. Some studies compare exam scores from different cohorts of students (Paschal, 2002; Knight and Wood, 2005; Freeman et al., 2007), whereas others compare how the same students performed on questions based on material covered with or without clickers (Bunce et al., 2006), or with different levels of clicker use (Suchman et al., 2006; Preszler et al., 2007). Other variables include the effects of different class sizes, different instruc-

tors (Suchman *et al.*, 2006), synergy with other active-learning techniques introduced at the same time as clickers (Knight and Wood, 2005; Freeman *et al.*, 2007), differences in course level and instructor experience with clicker technology (Preszler *et al.*, 2007), and even the effects of outside events on the comparison of different cohorts of students (Paschal, 2002). However, it is clear that when learning gains are seen, it is when clickers are used to enhance active learning (Judson and Sawada, 2002; Caldwell, 2007).

Many different teaching techniques can be used with the aid of clickers. Two techniques we used were "just in time teaching" (JiTT) and "peer instruction." JiTT involves the use of the Internet to give students a "warmup" assignment. The instructor uses the online responses of the students to design the lecture to more thoroughly discuss concepts that were difficult or that caused misconceptions among the students. To reinforce particular concepts, students partake in cooperative learning exercises in class (Novak *et al.*, 1999; Marrs and Novak, 2004). Peer instruction calls for students to answer a question on their own, record a response, and then attempt to convince their peers that their answer is correct before reporting the final answer to the instructor (Mazur, 1997; Crouch and Mazur, 2001). This type of instruction has been shown to improve student learning in several studies (Rao and DiCarlo, 2000; Nichol and Boyle, 2003; Duncan, 2005; Knight and Wood, 2005; Smith *et al.*, 2005).

We were interested in learning how teaching methods associated with using clickers might benefit student learning in different types of biology classes. Specifically, we compared an introductory course for nonmajors and a sophomore-level genetics course. In this article, we first describe a study that shows that the initial introduction of clickers into our respective courses did not have a significant effect on student performance on exams. We hypothesized that by fall 2006 we had sufficient experience using clickers to expect a positive effect of clickers on student performance. We performed a detailed comparison of student performance on exam questions based on concepts taught with clickers versus nonclicker-based exam questions for both courses. Last, we assessed student performance on 10 clicker-based and 10 nonclicker-based exam questions 4 mo after the end of the course to see whether teaching methods facilitated by clickers helped long-term retention of course material.

METHODS

Courses Using Clickers

Biological Foundations is a 100-level nonmajors biology course that students take to satisfy a "lab science" requirement at the University of Wisconsin–Whitewater (UWW). This course was taught in a large lecture hall. The data from two classes (87 and 107 students each) were combined for the analyses summarized below. Curran has taught Biological Foundations every semester since fall 2004 and had 4 yr of full-time teaching experience before the fall 2004 semester. Introduction to Genetics is a 200-level course that is required of all biology majors. This course has a smaller class size (46 students in fall 2006); therefore, it is taught in a smaller lecture hall than Biological Foundations. Crossgrove has taught genetics at UWW every year since fall 2004 and had 6.5 yr of full-time teaching experience before fall 2004. Before first using clickers in fall 2005, Curran and Crossgrove participated in several on-campus workshops during summer 2005 and attended a daylong workshop at

UW–Madison to learn about the technology and educational practices associated with using clickers. Curran taught the nonmajors biology course without clickers in fall 2004, spring 2005, and fall 2005, and with clickers in fall 2005, spring 2006, and fall 2006. Crossgrove taught genetics in fall 2004 and spring 2005 without clickers and taught spring 2006 and fall 2006 with clickers.

Student Response System Technology

UWW adopted TurningPoint (Turning Technologies, Youngstown, OH) with radiofrequency receivers as the supported clicker technology. A review of clicker devices, including TurningPoint, is available in Barber and Njus (2007). Students can buy their clickers through the bookstore, and they must register their clickers by using a website created by the UWW Learning Technology Center (LTC). No fee is assessed for registration. Both instructors allowed about 2 wk for the students to register their clickers.

TurningPoint works directly through PowerPoint (Microsoft, Redmond, WA), so it is very simple to use during a lecture. Still, some technical difficulties were encountered in using the software and hardware (Supplemental Material A). Both instructors included clickers as a participation grade, and they did not grade the students on their actual performance on the clicker questions. Curran had two large sections of students and processed clicker points after each lecture. The total points per student were then uploaded to the class website (Desire2Learn [D2L], Kitchener, ON, Canada) before every exam so that students could check their progress and make sure that the scores calculated were correct. Crossgrove had fewer students and uploaded the clicker points to D2L approximately every other week.

Teaching Techniques Used with the Aid of Clickers

During the fall 2006 study, both instructors normally asked between two and eight clicker questions per lecture. In the nonmajors biology course, a clicker question based on a concept covered from the previous lecture was made available ≈5 min before the start of class. During lecture, clickers were used to go over warmup questions (JiTT) that students had previously answered online. The student responses to the online questions (in which students had to rely on the reading for their answers) were then compared with their clicker responses after discussing the concept in lecture. Clicker questions not based on warmups were also included in each lecture. When fewer than 70% of the students selected the correct answer for a question, peer instruction was used, followed by repolling the question. If the correct answer was not obtained, the instructor would try to explain the concept again. The subsequent lecture would sometimes include a similar question to test whether the students better understood the concept.

Clickers were also used to review for exams using a "Jeopardy" style game (Supplemental Material B and C).

The genetics course was taught using a "problem-based" approach. Clickers were used primarily to assess whether students understood the material by having them do genetics problems in class. Time spent on particular concepts was adjusted based on student responses. Classes usually started with a clicker question. Often, the question had been given at the end of the previous class or was a homework question, but sometimes a new question was used. Problem-based clicker questions were also interspersed throughout the lectures. Students were given some time to work through the problem. If fewer than 70% of the students answered correctly, peer instruction was used. Clickers were used occasionally to help go over problem set answers on the day the problem sets were due (JiTT) and once to review for an exam using Jeopardy (Supplemental Material B and C).

Student Opinion Survey

A survey was administered at the end of the fall 2006 semester to students in Curran's nonmajors biology sections and Crossgrove's

genetics course to assess student responses to using clickers. The survey used a subset of questions from a longer survey prepared as part of a University of Wisconsin System Student Response System grant (full survey text is contained in Supplemental Material D). The survey included six MC questions asking for demographic information (e.g., age, gender, race, marital status, student status, and education level), two MC questions on previous and current clicker use, 14 MC questions on general computer use, 11 Likert-type questions on the experience of using clickers in the relevant course, and four open-ended questions on clicker use in the relevant course. Both positively (eight) and negatively (three) phrased Likert-type questions were included.

The survey was approved by the UWW Institutional Review Board for the Protection of Human Subjects (IRB Proposal Ratification W05609022X), and it was administered through a website set up by the UWW LTC. Students were informed of the survey in class, e-mail, and on the D2L course website. In class and at the onset of the survey, it was made clear that participation in the survey was voluntary and that only grouped data would be published. However, as an incentive, extra credit was offered for completing the survey. The LTC separated student identification from the survey results to preserve anonymity of the students. The students were given ≈1 wk to take the survey at the end of the semester; 185 of 194 Biological Foundations students (95%) and 44 of 46 Introduction to Genetics students (96%) participated in the survey.

Answers to the computer and clicker questions were converted to numerical values as follows. Computer question responses were coded as "always" = 4, "often" = 3, "sometimes" = 2, "rarely" = 1, and "never" = 0. Clicker question responses were coded as "strongly agree" = 5, "agree" = 4, "neutral" = 3, "disagree" = 2, and "strongly disagree" = 1. For each question, the average was calculated for all students in each course (nonmajors biology and genetics). Paired t tests were then performed to independently compare responses to the computer questions and to the clicker questions between the nonmajors biology and genetics students.

To independently compare the student responses to each question, chi square analysis using a contingency table was performed on each question in which the columns were nonmajors biology and genetics students and the rows were the actual number of students who selected each response. In some cases, very few students had selected a particular response, and response categories had to be combined to ensure that the expected values were at least five.

Exam Analysis

MC questions were used to analyze the effects of clicker use on student learning. All nonmajors biology exam questions were MC, whereas genetics exams varied from 48 to 70.5% MC. The student performance on each question was defined as the percentage of students answering each question correctly. For fall 2005, when the nonmajors introductory biology course was taught with one clicker section and one nonclicker section, the student performance on each of the final exam questions was compared between the two sections (same exam) by using a paired t test. The data had a less-than-optimal fit to a Normal distribution; therefore, we also performed a nonparametric test for two related samples (Wilcoxon signed ranks test) to verify the significance. For nonmajors biology, 10 assessment questions that are used by all instructors were analyzed for all of the sections that Curran has taught, including five sections taught without clickers and five sections taught with clickers. Crossgrove similarly analyzed 10 questions that were on all of her genetics final exams, composed of two sections that did not use clickers and two sections that used clickers. The data did not fit a Normal distribution; so, the student performance on these questions was analyzed for each course using a Wilcoxon signed ranks test to compare the overall student performance on each question for students in courses that did not use clickers to students in courses that used clickers. The dependent variable was the student performance on each question, and the test looked at the effects of clicker use

(whether the course was taught with clickers or not). For fall 2006, each MC exam question was classified as either a "clicker" or "nonclicker" question, based on whether the question covered material that had been taught using clickers (Bunce *et al.*, 2006; Suchman *et al.*, 2006). Exam questions did not have to be identical to clicker questions used in class during the semester, but they had to cover the same material in a similar manner. All of the exam questions from both nonmajors biology (225 questions) and genetics (93 questions) were analyzed using an analysis of variance (ANOVA). The dependent variable was the student performance on each question, and the ANOVA looked at the effects of course (nonmajors biology or genetics) and clickers (clicker vs. nonclicker question).

Each MC question used in the fall 2006 course analysis was also classified according to Bloom's taxonomy (Bloom, 1956) as to whether the question assessed knowledge, comprehension, application, or analysis. Both instructors and an outside reviewer independently classified all questions. The instructors then went through the classifications together and reached a consensus on all questions for which there was disagreement. An ANOVA was used to analyze whether there were any differences in student performance at the different cognitive levels of Bloom's taxonomy. The dependent variable was the student performance of each question. MC questions from all three semester exams and from the final exam were included for both nonmajors biology (225 questions) and genetics (93 questions). The ANOVA included the effect of clicker and nonclicker questions and the Bloom's taxonomy level of each question. There were 133 knowledge questions, 108 comprehension questions, 72 application questions, and five analysis questions. Because there were very few analysis questions, they were grouped with the application questions for statistical analysis.

Retention Assessment

To assess whether the clickers had an effect on how well students retained material after the completion of the course, we developed a postcourse "test" for each course, and we administered it to volunteers (students who had taken either Curran's nonmajors course or Crossgrove's genetics course in fall 2006). For each test, 20 questions were selected that had been on exams given during the course. Ten questions were on material covered using clickers, and 10 questions were on material covered without clickers. To focus on material that the students had understood well during the course, questions were limited to those in which the student performance during the semester was >70%. Paired clicker and nonclicker questions were matched reasonably closely in terms of student performance. Students were contacted by e-mail and asked to participate in a follow-up survey to be conducted on April 19 and 20, 2007, and they were offered free pizza as an incentive. The follow-up survey was approved by the UWW Institutional Review Board for the Protection of Human Subjects (IRB Proposal Modification W05609022X). The test results were anonymous (scan-tron), but we recorded which students participated in the test. Fourteen nonmajors biology students (of 194 possible participants) and 15 genetics students (of 46 possible participants) took the survey.

To analyze the retention test data, the student performance was calculated for each question on the survey. In addition, the student performance was calculated for each question given in fall 2006 for the students who took the retention test. The resulting data did not fit a Normal distribution; therefore, it was analyzed using a Wilcoxon signed ranks test. For each course, this test was used to compare student performance in fall 2006 and spring 2007. Next, fall 2006 and spring 2007 responses to clicker questions and fall 2006 and spring 2007 responses to nonclicker questions were compared.

All statistics calculations were performed using SPSS (SPSS, Chicago, IL).

RESULTS

Student Opinion Survey Responses

We compared the student opinions of clicker use in a nonmajors introductory biology course (Biological Foundations) and in a sophomore-level biology majors course (Introduction to Genetics), and we saw an overall difference in their opinions of the clickers ($t = -3.00$, d.f. = 10, $p = 0.013$) but no difference in computer use ($t = 0.333$, d.f. = 13, $p = 0.744$). Specifically, the nonmajors had a more positive opinion of clickers (average Likert score of 3.8) than the students in the majors course (average Likert score of 3.6). It should be noted that students in both courses had an overwhelmingly positive opinion of clicker use. Students in both courses felt the clickers helped them better understand concepts (79% agreed or strongly agreed) and increased class participation (80% disagreed or strongly disagreed with a negatively worded statement). In both courses, fewer students, but still a majority, felt the clickers stimulated interaction between students in the classroom (71% disagreed or strongly disagreed with a negatively worded statement),

made it easy to connect ideas together (69% agreed or strongly agreed), and motivated them to learn (59% agreed or strongly agreed). Students in both courses expressed dissatisfaction with the cost of the clickers, with nearly half (45%) agreeing or strongly agreeing that the clickers were too expensive for what they got out of it. About one-third (32%) overall were neutral on the subject, and only 21% disagreed or strongly disagreed.

Chi square analysis was performed on each question to determine the likely source of the differences between the two courses in their clicker responses. Although all of the questions discussed above showed no statistically significant differences by chi square analysis (Table 1), there were two questions in which students in the two courses differed significantly in their responses. The biggest difference occurred when students were asked whether they agreed with the statement that using clickers did not help them score higher on exams (Table 1). The nonmajors were much more likely than students in the genetics course to think that the clickers helped their exam performance. Specifically, 41% of genetics students agreed or strongly agreed with the nega-

Table 1. Means and chi square analysis of student responses to survey questions on clicker use comparing nonmajors- and majors-level biology courses

Question	Avg. response[a] (nonmajors/genetics)	Chi square analysis[b]	Combined answer categories[c]
Clickers made me feel involved in the course.	4.30/4.27	$\chi^2 = 5.46$, d.f. = 1, $p = 0.994$	Agree-strongly disagree
Using clickers helped me pay attention in class.	4.28/4.20	$\chi^2 = 1.94$, d.f. = 2, $p = 0.379$	Neutral-strongly disagree
Clickers allow me to better understand concepts.	4.10/3.91	$\chi^2 = 1.59$, d.f. = 2, $p = 0.451$	Neutral-strongly disagree
Using the clickers did not help me score higher on the exams.	**2.58/3.32**	$\boldsymbol{\chi^2 = 20.5}$, **d.f. = 2, $\boldsymbol{p < 0.001}$**	Strongly agree-agree; disagree-strongly disagree
The clicker was too expensive for what I got out of it.	3.36/3.60	$\chi^2 = 3.03$, d.f. = 3, p = 0.386	Disagree-strongly disagree
I would recommend that the instructor continue to use clickers.	**4.09/3.68**	$\boldsymbol{\chi^2 = 7.56}$, **d.f. = 2,** $p = $ **0.023**	Neutral-strongly disagree
Clickers did not stimulate interaction with my classmates.	2.34/2.35	$\chi^2 = 1.39$, d.f. = 3, $p = 0.708$	Strongly agree-agree
Clickers helped me get instant feedback on what I knew and didn't know.	4.32/4.39	$\chi^2 = 0.01$, d.f. = 1, $p = 0.92$	Agree-strongly disagree
The clickers used in this course motivated me to learn.	3.64/3.45	$\chi^2 = 1.84$, d.f. = 3, $p = 0.606$	Disagree-strongly disagree
The clickers made it easy to connect ideas together.	3.79/3.45	$\chi^2 = 5.82$, d.f. = 2, $p = 0.055$	Neutral-strongly disagree
The clickers did not increase my participation in class.	2.00/2.11	$\chi^2 = 544$, d.f. = 2, $p = 0.762$	Strongly agree-neutral

[a] Column 2 shows the average response for students in a nonmajors biology course (BIO 120: Biological Foundations) and a majors level course (BIO 251: Introduction to Genetics) where the average was calculated after setting all strongly agree answers to 5, agree to 4, neutral to 3, disagree to 2, and strongly disagree to 1.

[b] Column 3 shows the results of chi square analysis using a contingency table in which the number of students answering strongly agree, agree, neutral, disagree, or strongly disagree were analyzed for BIO 120 compared with BIO 251. The p values ≤ 0.05 are considered significant and are indicated in bold.

[c] For most questions, some of the possible answers were selected by so few individuals that the expected values for that answer were less than the required five needed to use chi square analysis. In those cases, answer categories were combined until the expected value was at least five. The answer categories that were combined in the analysis are indicated in column 4. Dashes are used to represent inclusive categories (i.e. neutral-strongly disagree means that the number of students answering neutral, disagree, and strongly disagree were combined), whereas semicolons separate different groups (i.e., strongly agree-agree; disagree-strongly disagree means that the number of students selecting strongly agree and agree were combined into one category and the number of students selecting disagree and strongly disagree were combined in another category).

tively worded statement compared with only 18% of non-majors, whereas only 18% of genetics students disagreed or strongly disagreed, compared with 55% of nonmajors. Another difference was observed when students were asked whether they agreed with recommending that the instructor continue to use clickers (Table 1). The majority of students in both courses agreed with the statement, but 81% of nonmajors agreed or strongly agreed compared with 64% of genetics students. Similar percentages of students disagreed or strongly disagreed with the statement (11%, majors; 8%, nonmajors). More genetics students (25%) were neutral than nonmajors (10%), and 1% of nonmajors did not give an answer.

Comparing the answers we obtained from open-ended questioning of the students again revealed interesting similarities and differences in responses of students in nonmajors and majors courses. When asked what they liked best about their experience using the clickers, the genetics students (34%) mentioned valuing instant feedback on their understanding of concepts more than the nonmajors (17%). In contrast, the nonmajors valued the increased participation in lecture more than the genetics students (31 vs. 20%, respectively). The genetics students placed more value on the instructor knowing how they were doing and adjusting the lecture pace, and on the influence the clickers had on class discussion. One student commented ". . . I think the best thing about the use of the clickers is that it lets the instructor understand how well the students are understanding the material, which results in further discussion and clarification of the subject." The genetics students also liked to know how they were doing with respect to their classmates. A student commented, ". . . I like to know that when it comes to some information that I am not the only one that doesn't know what I am doing." The nonmajors indicated that the clickers were fun (14%) and that polling of answers was anonymous (13%). One student commented, "I got dependent on them, there was a day that the clicker questions weren't working and that was probably the longest and most boring class ever!" Another said, "I felt a lot more involved in class."

We also asked whether the students experienced any barriers in their use of clickers. Although many students said they had no barriers (nonmajors, 25%; genetics students 20%), technical difficulties (31% nonmajors and 25% genetics students), expense (27% genetics students and 13% nonmajors), and remembering the clicker (12% nonmajors and 4% genetics students) were cited as barriers.

Analysis of Overall Student Performance on Exams

In fall 2005, Curran taught two sections of nonmajors biology, one section using clickers, and the other section without clickers. Each section was given the same exams, and the average number of students who answered each final exam question correctly (72 questions) was calculated for each section (Figure 1). There was no significant difference between the two sections ($t = 0.333$, d.f. $= 71$, $p = 0.740$). Next, we analyzed the influence of clickers on student performance on exams by evaluating selected final exam questions given during the four to five semesters we have taught each of our courses. Our department requires that 10 assessment questions be added to every final exam for nonmajors. Each question includes a paragraph of information on three dif-
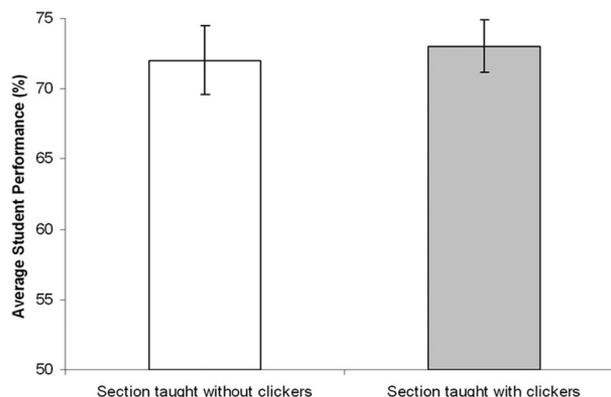

Figure 1. Comparison of final exam performance in courses taught with and without clickers. The bars represent the overall mean student performance on all final exam questions for a section of nonmajors biology taught with clickers compared with a section taught without clickers in the same semester (fall 2005). Both sections were given identical final exams. The error bars represent SE.

ferent topics and three to four questions about the paragraph. The three topics covered are the scientific method, genetics, and evolution. Student performance on these 10 assessment questions was compared between classes taught without clickers (fall 2004, spring 2005, and fall 2005; total of 463 students) and classes taught with clickers (fall 2005, spring 2006, and fall 2006; total of 425 students). The data did not fit a Normal distribution; so, we used a nonparametric test (Wilcoxon signed ranks test) to analyze the data. We found no significant difference in student performance between classes taught with clickers and classes that did not use clickers (Figure 2; Wilcoxon $z = -0.652$, $p = 0.515$).

Crossgrove taught genetics without clickers in fall 2004 and spring 2005 (total of 87 students). She taught the class with clickers in spring 2006 and fall 2006 (total of 96 students). She also analyzed 10 questions that were the same
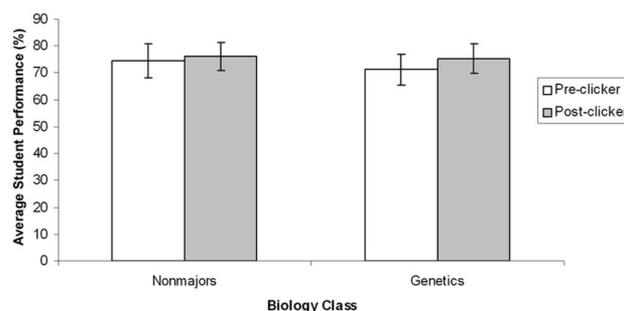

Figure 2. Comparison of student performance in courses taught with and without clickers from fall 2004 to fall 2006. The bars represent the overall mean student performance on 10 final exam questions given each semester in nonmajors biology and genetics (questions unique to each course). The empty bars represent the overall mean student performance (weighted average) in five sections of nonmajors biology (N = 463) and two sections of genetics (N = 87) that did not use clickers. The filled bars represent the overall mean student performance (weighted average) in five sections of nonmajors biology (N = 425) and two sections of genetics (N = 96) that did use clickers. The error bars represent SE.

between all semesters. There was no significant difference in student performance between classes taught with clickers or without clickers (Figure 2; Wilcoxon $z = -0.968$, $p = 0.333$)

Evaluation of Classes Taught with Clickers in Fall 2006

In fall 2006, we decided to see whether student performance increased when clickers are used primarily for comprehension, application, and analysis questions and when they are implemented by instructors experienced with the technology. We compared exam questions that were based on concepts covered using clicker questions with exam questions that were not covered with the aid of clickers (Bunce *et al.*, 2006; Suchman *et al.*, 2006). When all MC questions for both courses were analyzed by two-way ANOVA, there was a significant difference in how the questions were answered ($F = 3.465$, d.f. $= 3$, $p = 0.017$). There was a significant difference between the answers to clicker-based questions versus nonclicker-based questions ($F = 7.300$, d.f. $= 1$, $p = 0.007$). There was no difference in student performance between questions from the nonmajors course and questions from the genetics course ($F = 0.476$, d.f. $= 1$, $p = 0.491$), and there was no statistical interaction between the course with which the questions were associated and whether the questions were based on concepts taught with or without clickers ($F = 0.201$, d.f. $= 1$, $p = .655$). A direct comparison of the mean student performance ($\pm$SE) shows that the students performed better on concepts taught with clickers compared with concepts taught without clickers for both the nonmajors biology class (clicker, $74.1 \pm 1.7\%$; nonclicker, $66.5 \pm 1.8\%$) and the genetics class (clicker, $71.4 \pm 2.6\%$; nonclicker, $65.9 \pm 3.0\%$) (Figure 3).

We asked a variety of questions on our exams. When the exam questions were analyzed using Bloom's taxonomy, we found that ~41.8% of the questions were knowledge based (BT1), 34% of the questions were comprehension (BT2), and 24.2% were application or analysis (BT3/4). We were interested in analyzing whether there was any difference in how the students performed on different level questions. We categorized each exam question from fall 2006 for both classes as knowledge, comprehension, application, or analysis. We had very few analysis questions; therefore, application and analysis were combined. A two-way ANOVA was performed to compare the average number of questions answered correctly with the category of question asked. Again, the analysis showed that there was a statistical significance in the data ($F = 2.733$, d.f. $= 5$, $p = 0.02$), but it was due to whether clickers were used or not ($F = 11.569$, d.f $= 1$, $p = 0.001$) and not due to the level of question asked ($F = 0.464$, d.f. $= 2$, $p = 0.629$). This suggests that in general students performed better on any of the three types of questions (knowledge, comprehension, and application/analysis) when the concepts being tested were taught using clickers.
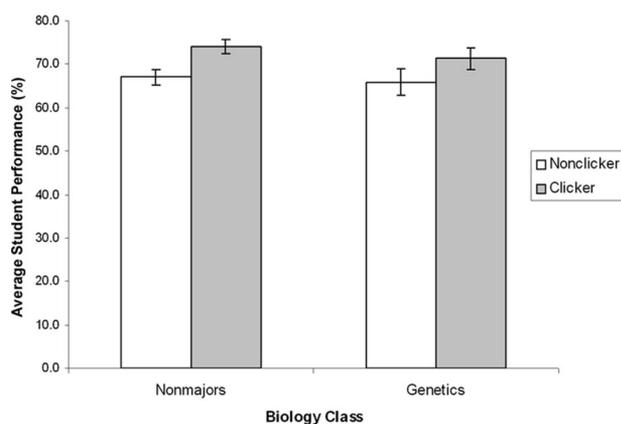
Improved Retention of Concepts with Clickers

To test whether students remembered concepts better over the long term when using teaching techniques associated with clickers, we first analyzed the exams from each course and picked clicker and nonclicker questions from each exam that had a 70% difficulty or higher (range from 70 to 100%). We assembled an assessment test of 20 questions (10 clicker-based and 10 nonclicker-based questions). Each clicker question was matched to a question of similar difficulty that was a nonclicker-based question. For the nonmajors course, the makeup of the assessment was roughly equal when analyzing the questions using Bloom's taxonomy: knowledge (three clicker, four nonclicker), comprehension (four clicker, three nonclicker), application (three clicker, three nonclicker). In contrast, for the genetics course the clicker questions consisted of comprehension (four) and application (six), whereas the nonclicker questions included knowledge (six), comprehension (two), application (one), and analysis (one) questions. Students were asked to come in and take this assessment test on April 19 and 20, 2007 (~4 mo after taking their final exam). Only 14 nonmajors and 15 genetics students participated in the postcourse assessment test. Each student took the test anonymously, but we recorded who took the test. This information was used to calculate student performance for fall 2006 on each of the assessment questions to accurately reflect the retention of knowledge for this subset of students.

Not surprisingly, students performed better on the assessment questions when taken during the course (fall 2006) compared with 4 mo later (spring 2007). For students in the nonmajors biology course, the average performance on clicker questions dropped from $88.3 \pm 5.1\%$ (SE) in fall 2006 to $83.1 \pm 5.1\%$ in spring 2007 (Figure 4). On nonclicker questions, the average performance dropped from $92.1 \pm 3.3\%$ in fall 2006 to $75.4 \pm 4.6\%$ in spring 2007. Due to small sample size and data that did not fit a Normal distribution, we used a nonparametric test to analyze the data. There was a significant difference between the fall 2006 and spring 2007 responses (Wilcoxon $z = -2.506$, $p = 0.012$), but when the clicker and nonclicker questions were analyzed separately, only the difference between fall and spring for the nonclicker questions was significant (nonclickers: Wilcoxon $z = -2.668$, $p = 0.008$; clickers: Wilcoxon $z = -0.892$, $p = 0.373$). The 15 genetics students tested had an average performance of $87.3 \pm 3.1\%$ in fall 2006 and $60.0 \pm 4.6\%$ in spring 2007 on


Figure 3. Comparison of student performance on clicker-and nonclicker-based questions in fall 2006. The bars represent the overall mean student performance on MC exam questions based on material covered using clickers (filled bars) or material covered without use of clickers (open bars) in nonmajors biology and genetics in fall 2006. The error bars represent SE.
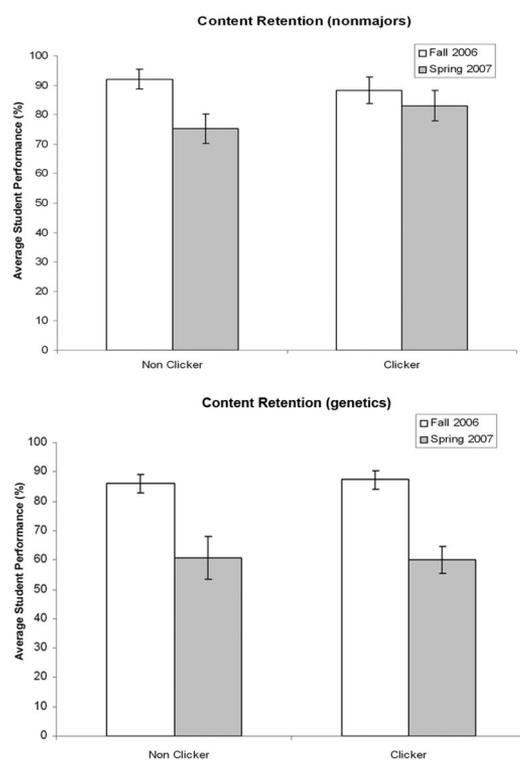
Figure 4. Retention of content 4 mo after completion of nonmajors biology and genetics courses. The bars represent the overall mean student performance on 10 MC questions based on material covered using clickers and 10 MC questions based on material covered without clickers. Open bars show the mean student performance on the questions given during the fall 2006 semester. Filled bars show the mean student performance on the same questions given ~4 mo after the end of the course. (A) Student performance on questions given during and after fall 2006 nonmajors biology course. (B) Student performance on questions given during and after fall 2006 genetics course. Error bars represent SE.

clicker-based questions, compared with 86 ± 3.1% average performance in fall 2006 and 60.7 ± 7.2% in spring 2007 for nonclicker questions. There was a significant difference between the fall 2006 and spring 2007 responses (Wilcoxon z = −3.826, p < 0.001). When clicker and nonclicker questions were analyzed individually, the difference between fall 2006 and spring 2007 was significant for both clicker and nonclicker questions (clicker: Wilcoxon z = −2.805, p = 0.005; nonclicker: Wilcoxon z = −2.668; p = 0.008).

DISCUSSION

In this article, we showed that students in both a nonmajors biology course and a genetics course for majors valued the use of clickers in the classroom. The use of clickers and the teaching techniques associated with them increased student performance on clicker-based exam questions compared with nonclicker-based exam questions. In the nonmajors biology course, students better retained knowledge from clicker-based exam questions compared with nonclicker-based exam questions.

Student Opinion

As in previous studies (Preszler *et al.*, 2007; reviewed in Judson and Sawada, 2002; Caldwell, 2007), students generally had a positive opinion about clicker use in the classroom, and many of the opinions did not differ significantly between the two classes. The two classes differed in their opinion on how the clickers affected their exam scores (Table 1). The nonmajors biology students thought that clickers helped their exam performance, whereas genetics students were not as convinced. Our analysis of student performance on clicker- and nonclicker-based exam questions showed that students performed better on clicker-based questions compared with nonclicker-based questions (Figure 3). However, on closer inspection the data show that there is more of a difference between student performance on clicker- and nonclicker-based questions in the nonmajors biology course than in the genetics course (7.6 and 5.5 percentage point differences, respectively), although it is not statistically significant. Another difference between the two classes was whether the students thought that the instructors should continue using clickers. The nonmajors biology students had a more positive response, which is similar to the findings of Preszler *et al.* (2007). Therefore, in general, students in introductory courses are more receptive to using clickers. We suspect that this is because the clicker use helps engage these students in the classroom and provides practice answering higher-order questions (e.g., comprehension, application, and analysis). This helps these students feel involved in a class that consists of a large number of people, helps them interact with their peers, and aids in connecting ideas together and understanding concepts. Our data suggest that students in both courses experience these benefits when clickers are used, but it is possible that the students who are in a larger class, or students who are not familiar with the subject, or both, may appreciate this benefit more.

Clickers and Student Performance

As in some previous studies, we found that clickers had no overall effect on exam performance when comparing separate cohorts of students (between classes and semesters) (Figures 1 and 2; Bessler and Nisbet, 1971; Casanova, 1971; Paschal, 2002). There are some factors that may have influenced these results. For example, the courses taught with clickers displayed many of the growing pains associated with adapting lectures and lecture style to new technology. Another potential confounding factor is the fact that both instructors had already incorporated active learning to varying degrees into their courses before beginning to use clickers. It is well established that active learning can improve student performance (Hake, 1998, 2002; Meltzer and Manivannan, 2002; Udovic et al., 2002; Smith *et al.*, 2005). Before incorporating clickers, Curran used JiTT in the nonmajors course and Crossgrove used a problem-based approach in the genetics course in which students worked together on genetics problems during class. As we became more comfortable with clicker use, a concerted effort was made (fall 2006) to use clickers for comprehension, application, and analysis questions (Bloom, 1956) and to more rigorously incorporate peer instruction (Mazur, 1997; Duncan, 2005). In addition, we decided to test for influences of clickers on student performance within the same cohort of

students. We thought that this type of analysis might tease out the influences of clickers on student learning while controlling for variation between the caliber of students present in different sections of a course or between semesters. Using this approach, we saw a statistically significant effect on the students' performance (Figure 3). A similar study by Bunce *et al.* (2006) found no significant difference between clicker-based questions and nonclicker-based MC questions, although the average was higher for the clicker-based questions. In addition, we found that gains occurred irrespective of the type of question asked (e.g., knowledge, comprehension, and application/analysis). In contrast, Slain *et al.* (2004) found that student performance only increased on clicker-based questions that were analytical.

We think that engaging the students in the classroom through clicker questions and using them to concentrate on higher-order questions (e.g., comprehension, application, and analysis) in class gives students good practice in critical-thinking skills (Freeman *et al.*, 2007). This practice may have an overall effect of helping our weaker students perform at a higher level. One interesting observation we made was that we consistently saw that the SD on clicker-based questions was smaller than on nonclicker-based questions. Similar observations have been reported by others (Poulis *et al.*, 1998), with the interpretation that students are able to answer clicker-based questions more consistently than the nonclicker-based questions. Also, even though the difference in student performance seen between clicker- and nonclicker-based questions was small (Figure 3), the increase in performance was the difference between a "D" and a "C".

Long-Term Retention of Concepts

The use of clicker-based teaching techniques improved the long-term retention of material for the nonmajors biology class but not for the genetics class. One of the unanswered questions in clicker literature today is if this type of teaching technique can help students remember concepts over the long term. We tested students ~4 mo after they had taken their final exam for the course. Very few students (14 nonmajors biology students and 15 genetics students of a possible total of 240 students) participated (even with the promise of free food). For the nonmajors biology students, the location of the exam may have been a factor in the low turnout. We held the assessment exam in our science building instead of the student center or another more central location. Perhaps if the location had been more accessible, we may have obtained more participants. Nonparametric analysis of the data from the nonmajors biology class showed no significant change in student performance on clicker-based questions between fall and spring. This suggests that the students were performing almost as well on these questions 4 mo later as they did when taking the course. There was a statistically significant difference in student performance between fall and spring for nonclicker-based questions, indicating that students performed less well 4 mo later than when they originally were tested on the material.

The genetics students did not show any difference in their retention of material for clicker-and nonclicker-based questions, but there was more variation in the responses in spring 2007 to nonclicker questions than clicker questions. One explanation for this could be that the types of questions

that were used were not evenly matched when comparing the performance on clicker-and nonclicker-based questions. This was a consequence of picking questions based on how well the students had performed when given the questions on exams during fall 2006. Because clickers were used primarily to work on genetics problems, the pool of clicker questions tended to be mostly application questions with some comprehension questions. In contrast, nonclicker-based questions were primarily "lower-order" knowledge or comprehension questions.

Most instructors hope that the knowledge they impart to their students will be remembered. Semb and Ellis (1994) provide a careful review of research findings associated with retention of knowledge taught in the classroom. They conclude that instructional strategies that increase the level of "original learning" lead to greater retention of the knowledge over the long term. It is possible that the instructive techniques used while using clickers (peer instruction, JiTT, and practice with higher-order questions) may enhance the original learning of our students and ultimately lead to greater retention of the concepts covered using those techniques. Naceur and Schiefele (2005), who tested retention of knowledge over 3 wk, suggest that students' interest in a subject affects their long-term retention. It is possible that the use of clicker-based teaching methods and the use of a game-like review of material (Supplementary Material B) may make the course more interesting and could therefore affect long-term retention of concepts.

Another hypothesis for the increased retention we saw in our study is that it is due to the feedback students get through JiTT, peer instruction, and Jeopardy that is executed through the use of clickers. These techniques inform the students on their level of understanding of the material and help them gauge how much more they need to study. It is interesting to note that the genetics course did not regularly use JiTT with clickers and only used Jeopardy to review for one exam, so those students did not get the same level of feedback as in the nonmajors course. Cooperative testing and quizzing has been found to influence retention of material for up to 4 wk (Cortright *et al.*, 2003; Zeilik and Morris, 2004). The effects of instant feedback on student retention have also been studied. Most of the studies done on the effects of feedback on memory have tested short-term memory. Guthrie (1971) showed that feedback helped learning when the answer was wrong (retention test given immediately after information was learned). A more recent study suggests that figuring out why answers are wrong helps to break misconceptions (Tanner and Allen, 2005). However, other research is less clear about the benefits of feedback (Kulik and Kulik, 1988).

In conclusion, clickers are a useful tool for increasing student and instructor satisfaction and student learning. The clickers should be incorporated as part of a general strategy of active learning. Increases seen in student performance are likely due to the increase in active learning facilitated by clickers. Therefore, gains may be more subtle if active learning is already used. Interestingly, long-term retention of information may be enhanced with active-learning techniques executed with the help of clickers. We are interested in whether long-term retention of concepts (beyond the scope of the course) can be influenced by active-learning

techniques and hope that these results will spur more research in this area.

REFERENCES

Barber, M., and Njus, D. (2007). Clicker evolution: seeking intelligent design. CBE Life Sci. Educ. *6*, 1–8.

Bessler, W. C., and Nisbet, J. J. (1971). The use of an electronic response system in teaching biology. Sci. Educ. *55*, 275–284.

Bloom B. S. (1956). Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain, New York: David McKay Co. Inc.

Bunce, D. M., Van den Plas, J. R., and Havanki, K. L. (2006). Comparing the effectiveness on student achievement of a student response system versus online WebCT quizzes. J. Chem. Educ. *83*, 488–493.

Caldwell, J. (2007). Clickers in the large classroom: current research and best-practice tips. CBE Life Sci. Educ. *6*, 9–20.

Casanova, J. (1971). An instructional experiment in organic chemistry. J. Chem. Educ. *48*, 453–455.

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., and DiCarlo, S. E. (2003). Student retention of course content is improved by collaborative-group testing. Adv. Physiol. Educ. *27*, 102–108.

Crouch, C. H., and Mazur, E. (2001). Peer instruction: ten years of experience and results. Am. J. Phys. *69*, 970–977.

Duncan, D. (2005). Clickers in the Classroom: How to Enhance Science Teaching Using Classroom Response Systems, New York: Addison-Wesley and Benjamin Cummings.

Freeman, S., O'Connor, E., Parks, J. W., Cunningham, M., Hurley, D., Haak, D., Dirks, C., and Wenderoth, M. P. (2007). Prescribed active learning increases performance in introductory biology. CBE Life Sci. Educ. *6*, 132–139.

Guthrie, J. T. (1971). Feedback and sentence learning. J. Verbal Learn. Verbal Behav. *10*, 23–28.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: a six-thousand student survey of mechanics test data for introductory physics courses. Am. J. Phys. *66*, 64–74.

Hake, R. (2002). Lessons from the physics education reform effort. Conserv. Ecol. *5*, 28.

Judson, E., and Sawada, D. (2002). Learning from the past and present: electronic response systems in college lecture halls. J. Comput. Math. Sci. Teach. *21*, 167–181.

Knight, J. K., and Wood, W. B. (2005). Teaching more by lecturing less. Cell Biol. Educ. *4*, 298–310.

Kulik, J. A., and Kulik, C.-L.C. (1988) Timing of feedback and verbal learning. Rev. Educ. Res. 38, 79–97.

Marrs, K. A., and Novak, G. (2004). Just-in-time-teaching in biology: creating an active learner classroom using the Internet. Cell Biol. Educ. *3*, 49–61.

Mazur, E. (1997). Peer Instruction: A User's Manual, Upper Saddle River, NJ: Prentice Hall.

Meltzer, D. E., and Manivannan, K. (2002). Transforming the lecture-hall environment: the fully interactive physics lecture. Am. J. Phys. *70*, 639–654.

Naceur, A., and Schiefele, U. (2005). Motivation and learning–the role of interest in construction of representation of text and long-term retention: inter- and intraindividual analyses. Eur. J. Psychol. Educ. *20*, 155–170.

Nichol, D. J., and Boyle, J. T. (2003). Peer instruction versus class-wide discussion in large classes: a comparison of two interaction methods in the wired classroom. Stud. Higher Educ. *28*, 457–473.

Novak, G., Patterson, E. T., Gavrin, A. D., and Christian, W. (1999). Just-In-Time Teaching: Blending Active Learning with Web Technology, Upper Saddle River, NJ: Prentice Hall.

Paschal, C. B. (2002). Formative assessment in physiology teaching using a wireless classroom communication system. Adv. Physiol. Educ. *26*, 299–308.

Poulis, J., Massen, C., Robens, E., and Gilbert, M. (1998). Physics lecturing with audience paced feedback. Am. J. Phys. *66*, 439–441.

Preszler, R., Dawe, A., Shuster, C. B., and Shuster, M. (2007). Assessment of the effects of student response systems on student learning and attitudes over a broad range of biology courses. CBE Life Sci. Educ. *6*, 29–41.

Rao, S. P., and DiCarlo, S. E. (2000). Peer instruction improves performance on quizzes. Adv. Physiol. Educ. *24*, 51–55.

Semb, G. B., and Ellis, J. A. (1994). Knowledge taught in school: what is remembered? Rev. Educ. Res. *64*, 253–286.

Slain, D., Abate, M., Hodges, B. M., Stamatakis, M. K., and Wolak, S. (2004). An interactive response system to promote active learning in the doctor of pharmacy curriculum. Am. J. Pharm. Educ. *68*, 1–9.

Smith, A. C., Stewart, R., Shields, P., Hayes-Klosteridis, J., Robinson, P., and Yuan, R. (2005). Introductory biology courses: a framework to support active learning in large enrollment introductory science courses. Cell Biol. Educ. *4*, 143–156.

Suchman, E., Uchiyama, K., Smith, R., and Bender, K. (2006). Evaluating the impact of a classroom response system in a microbiology course. Microbiol. Educ. *7*, 3–11.

Tanner, K., and Allen, D. (2005). Approaches to biology teaching and learning: understanding the wrong answers—teaching toward conceptual change. Cell Biol. Educ. *4*, 112–117.

Udovic, D., Morris, D., Dickman, A., Postlethwait, J., and Wetherwax, P. (2002). Workshop biology: demonstrating the effectiveness of active learning in an introductory biology class. Bioscience *52*, 272–281.

Zeilik, M., and Morris, V. J. (2004). The impact of cooperative quizzes in a large introductory astronomy course for non-science majors. Astron. Educ. Rev. *3*, 51–61.