Article

Using Clickers to Facilitate Development of Problem-Solving Skills

Aime A. Levesque

Department of Biology, University of Hartford, West Hartford, CT 06117

Classroom response systems, or clickers, have become pedagogical staples of the undergraduate science curriculum at many universities. In this study, the effectiveness of clickers in promoting problem-solving skills in a genetics class was investigated. Students were presented with problems requiring application of concepts covered in lecture and were polled for the correct answer. A histogram of class responses was displayed, and students were encouraged to discuss the problem, which enabled them to better understand the correct answer. Students were then presented with a similar problem and were again polled. My results indicate that those students who were initially unable to solve the problem were then able to figure out how to solve similar types of problems through a combination of trial and error and class discussion. This was reflected in student performance on exams, where there was a statistically significant positive correlation between grades and the percentage of clicker questions answered. Interestingly, there was no clear correlation between exam grades and the percentage of clicker questions answered correctly. These results suggest that students who attempt to solve problems in class are better equipped to solve problems on exams.

INTRODUCTION

The most important skill an undergraduate science student can develop is the ability to solve problems. To solve a novel problem, the student must have a firm understanding of the scientific concepts underlying the problem, as well as the ability to think critically to analyze the concepts from different perspectives. It should come as no surprise that the development of this ability requires practice. If a student is never challenged to apply critical-thinking skills to analyze novel problems, he or she will never develop the ability to solve problems independently.

A major challenge instructors face in the typical undergraduate classroom is how to encourage all students to practice

the application of critical thinking to solving problems during lectures. It has been shown that student learning outcomes in biology classes, particularly in the area of conceptual understanding, increase when students are encouraged to engage in cooperative problem solving during lectures, through either small-group peer instruction or other interactive group activities (Mazur, 1997; Crouch and Mazur, 2001; Tanner *et al.*, 2003; Knight and Wood, 2005; Armstrong *et al.*, 2007; Crouch *et al.*, 2007; Preszler, 2009; Herreid, 2010). The value of peer instruction stems largely from the benefits of class discussion reinforcing concepts (Tanner, 2009). Previous studies have revealed that when a student explains a concept to others, it reinforces that student's understanding (Chi *et al.*, 1994; Coleman *et al.*, 1997; Coleman, 1998). Additionally, students often appreciate hearing an explanation from their peers, rather than from the instructor alone, because they can relate to the perspectives of other students more readily (Preszler, 2009). Thus, the increased involvement of each student inherent in small-group peer instruction provides the students with the ideal opportunity to practice explaining concepts to one another.

A potential complication of peer instruction is that weaker students may rely on stronger students to figure out the answer, rather than attempting to solve the problem independently. In this scenario, the weaker students may still be unable to think independently to solve homework and exam

Table 1. Class demographics in genetics for four consecutive semesters

Semester[a]	Number of students	Average GPA[b]	Biology	Chemistry/biology	Health sciences	Physical therapy	Other[c]
Spring 2008	36	3.01	39%	19%	28%	6%	8%
Fall 2008	39	3.19	10%	26%	33%	8%	23%
Spring 2009	30	2.91	47%	13%	27%	0%	13%
Fall 2009	33	3.13	16%	39%	18%	12%	15%

[a]Clickers were used in Spring and Fall 2009 but not in Spring and Fall 2008.

[b]This value is based on the GPA of each student as of the end of the preceding semester.

[c]Other majors consist primarily of the following: medical technology, radiological technology, respiratory therapy, biomedical engineering, and postbaccalaureate/premedical studies.

problems. To address this, instructors often couple peer instruction with classroom response systems, or clickers. Clickers have grown in popularity in recent years, largely because of their value in engaging all students during lectures, particularly in large classes (Draper and Brown, 2004; Caldwell, 2007; L.J. Collins, 2007; Cain and Robinson, 2008; J. Collins, 2008). Several studies have demonstrated that the use of clickers during lectures increases student performance on exams in undergraduate science classes (e.g., Preszler et al., 2007; Crossgrove and Curran, 2008; Reay et al., 2008). Clickers have also been used to promote development of clinical reasoning and decision-making skills among nursing students (DeBourgh, 2008; Russell et al., 2011). Thus, clickers may represent a valid tool to promote problem solving and critical thinking in all students during lectures.

One method for overcoming the influence of stronger students on weaker students is to combine independent problem-solving approaches with peer instruction. In this method, students are given an opportunity to attempt to solve a problem independently using clickers before they discuss the problem with their peers. In many studies, students vote on a question, observe the histogram of student responses (but not the correct answer), discuss the answer with their peers, and then revote on the same question (e.g., Allen and Tanner, 2002; Knight and Wood, 2005). In general, the number of students choosing the correct response on the revote increases. Unfortunately, students may be biased by seeing the most common response from the first vote and are more likely to simply choose the most common answer on the revote, as was illustrated in a previous study (Perez et al., 2010). To assess whether the increased number of correct responses on the revote is due to an increased understanding of the concept after peer discussion, it may be more effective to ask a similar follow-up question, rather than just revoting on the same question. In one study, students' abilities to answer a similar follow-up question increased, even when none of the students in a discussion group answered the first question correctly (Smith et al., 2009). In that particular study, student performance on exams was not analyzed, so it is not clear whether this improved understanding was reflected in an improved performance on exams.

The goal of the present study is to assess the effectiveness of using clickers to promote independent problem solving in a small class setting (~30 students). The study was carried out in a genetics course for majors during two semesters. The study design involved the independent solving of a novel problem using clickers, observation of the histogram of student responses, class and/or peer discussion, and a follow-up problem similar to the initial problem. The effectiveness of this approach in promoting problem-solving skills was assessed using correlation analysis. The results suggest that the use of clickers to solve problems during class correlates with student performance on exams stressing problem-solving and critical-thinking skills, regardless of whether the students answered the questions correctly during class.

METHODS

Course Background

Clickers were used in a sophomore-level genetics lecture course at the University of Hartford during two semesters: Spring 2009 and Fall 2009. The classes met 2 d per week for 1.25 h each day. In this study, two semesters in which clickers were used (Spring 2009 and Fall 2009) were analyzed, as were two semesters in which clickers were not used (Spring 2008 and Fall 2008). Class enrollments each semester were as follows: 36 students in Spring 2008, 39 in Fall 2008, 30 in Spring 2009, and 33 in Fall 2009 (Table 1). The distribution of majors varied each semester (Table 1), but the primary majors represented were biology, chemistry/biology, and health sciences.

To illustrate comparability of student populations in the four different semesters, the average class GPA was determined using each student's cumulative GPA as of the end of the last semester prior to taking genetics (Table 1). To compare the GPA values among the four semesters, Fisher's least significant difference post hoc test for pairwise differences was applied using PASW Statistics, version 18 (SPSS, Chicago, IL). The mean difference between two semesters was considered to be significant at a significance level of 0.05. The only two semesters for which values fell under this significance level were Fall 2008 and Spring 2009 ($p = 0.045$). This means that there was a statistically significant difference in the GPA values between these two semesters. However, because there was not a statistically significant difference between either of these two semesters and any of the other two semesters ($p > 0.05$), the slight difference between these two semesters is not substantive. Additionally, because the focus of this study is in the comparison between two semesters when clickers were not used (2008) and two semesters when clickers were used (2009), the fact that the average GPA by year did not differ significantly (3.11 in 2008 vs. 3.02 in 2009) suggests that the student populations are comparable.

A

a) A woman with type AB blood has a child with type A blood. Which of the following men definitely can not be the father?
1. A
2. B
3. AB
4. O
5. All of these are possible fathers

b) A woman with type A blood has a child with type O blood. Which of the following men definitely can not be the father?
1. A
2. B
3. AB
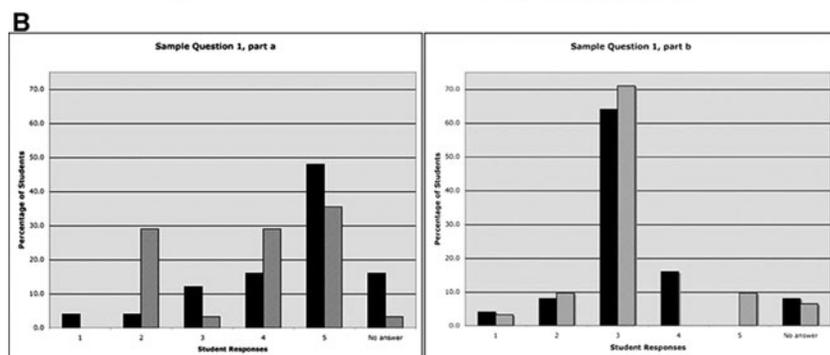4. O
5. All of these are possible fathers

B

Figure 1. Sample question 1 (A) and student responses (B). The correct response for question 1, part a, is 5; the correct response for question 1, part b, is 3. Responses from Spring 2009 are shown as black bars, and the responses from Fall 2009 are shown as gray bars. These data include only the students present in class at the time the problem was presented (25 in Spring 2009, 31 in Fall 2009).

The same instructor (A.A.L.) taught all four lecture sections, and the only difference in the format of the sections was the presence or absence of clickers. The format of the lectures consisted of segments of lecture covering concepts interspersed with work on sample problems. In addition to the in-class sample problems, homework problems were assigned in all four semesters. These included short problem sets (approximately two to three questions per chapter) that were collected and graded, as well as textbook practice problems (~10–20 questions per chapter) that were not collected. These homework assignments were given after class discussion of the relevant material. The problem sets contained similar types of problems all four semesters, but the specific questions were not identical. The same textbook practice problems were assigned all four semesters.

Clickers and Software

The clickers used in this study were TurningPoint from Turning Technologies (Youngstown, OH). The clickers were purchased using funds from a grant from the Emerging Technologies Pilot Program at the University of Hartford. Each student was assigned a specific clicker, and students collected their assigned clickers at the beginning of each class period and returned them at the end of each class period. A participant list was generated in the TurningPoint 1.1.4 program so that the responses of each student could be gathered for later analysis. Participant reports were generated in Turning Reports 1.1.4 software, and these responses were exported to Excel (Microsoft, Redmond, WA). Once all data for each student (clicker responses, attendance, and exam grades) were gathered in Excel, student names were removed from the spreadsheet to ensure anonymity.

Study Design

After a brief lecture on a concept, students were presented with a problem related to that concept and were given an opportunity to work independently to solve the problem without any prior discussion or explanation of how to solve the problem. For some problems, the clicker voting slide was shown concurrently with the problem, whereas for other problems the clicker voting slide was shown after students were given sufficient time to work on the problem. During the two semesters of this study in which clickers were used (Spring and Fall 2009), the same approach was used for each individual question to maintain consistency between semesters. The answers on the voting slides consisted of the correct answer along with several other common incorrect answers, which helped in identifying and addressing popular misconceptions and common mistakes (Tanner and Allen, 2005). Once approximately 75% of the class had voted, a 10-s countdown was displayed on the screen to give the rest of the students an opportunity to respond. The histogram of student responses was then immediately displayed on the screen, but the correct answer was not indicated. This was followed by period of discussion in which students explained why they chose their answers. In most instances, a whole-class discussion format was used, but small-group peer discussion was also used in some cases. A similar problem was then presented to the class, so that students would have a second opportunity to practice applying the same concepts to solving a similar problem. The histograms shown in this article were generated in Excel and are not the actual histograms that were shown in class (Figures 1B, 2C, 3C, and 4B). This was done in order to include the number of nonresponders for each question. Nonresponders are students who were present in class but did not answer the question. Absent students were not included among nonresponders.

In the two semesters when clickers were not used (Spring and Fall 2008), the same sample problems were presented to students in class, but students worked independently to solve the problems, and usually only a few students participated in discussing the correct answer. In these semesters, the problems were presented as open-ended questions, and students were not given multiple-choice answers as they were in the semesters when clickers were used.

Although clickers were used throughout the entire Spring and Fall 2009 semesters, this study focuses on the clicker questions used during the eight lectures in which material on the second exam was covered. This consisted of a total of 48 questions in Spring 2009 and 34 questions in Fall 2009. The number of clicker questions was smaller in Fall 2009, because

A

Coat color in Labrador retrievers is encoded by two separate gene loci. At the "B" locus the dominant B allele encodes black pigment and the recessive b allele encodes brown pigment. At the "E" locus, the dominant E allele is required for normal deposition of the black or brown pigment within the hair shaft, and the recessive e allele results in yellow coat color. Color is determined in the following manner:

B_E_ = black
bbE_ = brown
B_ee, bbee = yellow

Determine the **complete** <u>genotypes</u> of the **parents** for the following matings:

Phenotypes of Parents	**Phenotypes of Offspring**
a) Black X Brown	3/4 black; 1/4 yellow
b) Black X Yellow	1/2 yellow; 1/4 black; 1/4 brown
c) Black X Yellow	1/2 yellow; 3/8 black; 1/8 brown

B

What are the genotypes for cross a? | What are the genotypes for cross b? | What are the genotypes of cross c?

1. BbEe X bbEe | 1. BbEe X BbEe | 1. BbEe X BbEe
2. BBEe X bbee | 2. BbEe X bbee | 2. BbEe X bbee
3. BBEe X bbEe | 3. BbEe X bbEe | 3. BbEe X bbEe
4. BbEe X Bbee | 4. BbEE X bbee | 4. BbEe X Bbee
5. BbEE X bbEE | 5. BbEE X bbEE | 5. BbEE X bbEE
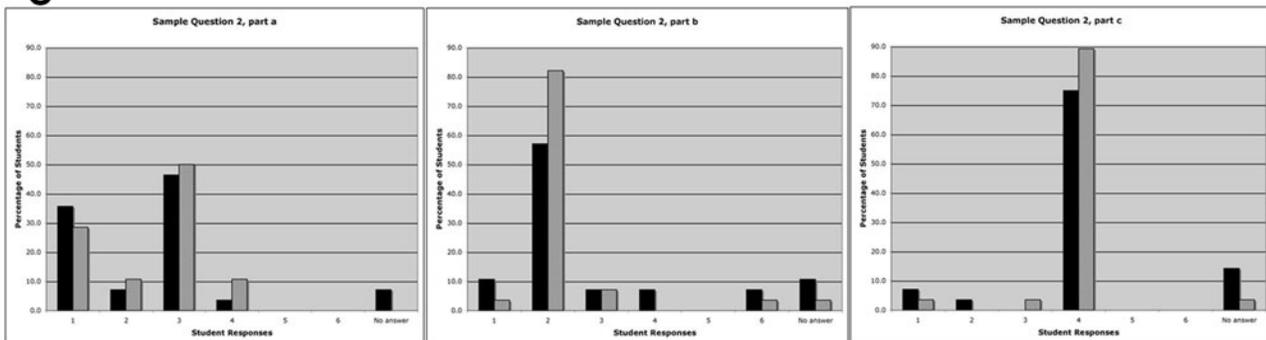6. None of these | 6. None of these | 6. None of these

C

Figure 2. Sample question 2 (A), clicker questions (B), and student responses (C). The correct response for question 2, part a, is 3; the correct response for question 2, part b, is 2; the correct response for question 2, part c, is 4. Responses from Spring 2009 are shown as black bars, and the responses from Fall 2009 are shown as gray bars. These data include only the students present in class at the time the problem was presented (28 in Spring 2009, 28 in Fall 2009).

the number of clicker questions used throughout the semester in Spring 2009 led to a major delay in the forward progress of the course. Some of the clicker questions for the Fall 2009 semester were cut, but the remaining questions were identical to those used in Spring 2009.

The second exam was chosen as the focus for this study for several reasons. The primary reason was to simplify and focus the scope of the study. Second, there is an adjustment period leading up to the first exam, in which students are acclimating to the use of clickers and to the format of the class. Clickers are not widely used at the University of Hartford, so for most students this was their first exposure to this technique. Third, the clicker questions used and the number of questions varied tremendously in the second half of the semester (corresponding to exams 3 and 4) between the Spring 2009 and Fall 2009 semesters, so any data generated

would not be comparable. This was not the case for the first half of the semester, when the questions used were far more consistent between the two semesters, although fewer questions were used in the Fall 2009 semester. Finally, the second exam covers sufficient problem-based concepts (multiple alleles, codominance, epistasis, sex-influenced traits, linkage, etc.) to serve as a valid assessment of problem-solving abilities.

Clicker Participation

For each student, the percent clicker participation was calculated based on the number of clicker questions the student answered out of the total questions (48 in Spring 2009, 34 in Fall 2009). Only nine of these questions are presented in this article. For each student, the percentage of correct answers

A

In foxgloves (*Digitalis purpurea*), flower coloration is influenced by two separate gene loci. The first gene determines the amount of pigment that is produced. The dominant D allele results in production of large amounts of pigment (dark purple), and the recessive d allele results in low amounts of pigment (light purple). The second gene affects pigment deposition in the petals. The dominant W allele prevents pigment deposition in the flowers (resulting in white flowers) whereas the recessive w allele allows normal pigment deposition in the flowers. Color in determined in the following manner:

D_W_ =white
ddW_ =white
D_ww =dark puple
ddww =light purple

For the following crosses, determine the genotypes of the parents.

Phenotypes of Parents	**Phenotypes of Offspring**
a) Light Purple X White	1/2 white; 1/4 dark purple; 1/4 light purple
b) White X Dark Purple	1/2 white; 1/4 dark purple; 1/4 light purple

B

a) What are the genotypes for cross a?
1. Ddww X ddWw
2. DdWw X ddww
3. ddww X DdWw
4. ddWw X Ddww
5. None of these

b) What are the genotypes for cross b?
1. Ddww X ddWw
2. DdWw X ddww
3. ddww X DdWw
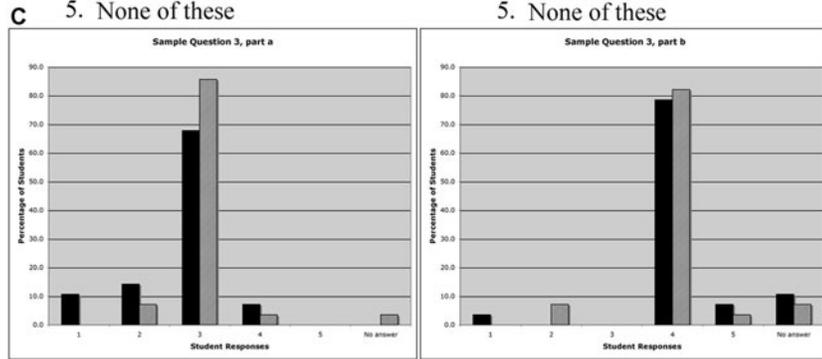4. ddWw X Ddww
5. None of these

C

Figure 3. Sample question 3 (A), clicker questions (B), and student responses (C). The correct response for question 3, part a, is 3; the correct response for question 3 part b, is 4. Responses from Spring 2009 are shown as black bars, and the responses from Fall 2009 are shown as gray bars. These data include only the students present in class at the time the problem was presented (28 in Spring 2009, 28 in Fall 2009).

was calculated based on the number of correct answers out of the total number of questions the student answered. For example, if a student in the Spring 2009 semester answered 40 questions, his or her clicker participation grade would be $(40/48) \times 100\% = 83\%$. If the student answered 38 questions correctly, his/her percent correct answers would be $(38/40) \times 100\% = 95\%$.

Clicker participation represented half of each student's class-participation grade, which constituted 5% of the total semester grade. The remaining half of the participation grade

A

a) If you cross a chinchilla rabbit with a chinchilla rabbit, and the offspring are 3/4 chinchilla and 1/4 albino, what were the genotypes of the parents?
1. $c^{ch}c$ X $c^{h}c^{h}$
2. $c^{ch}c^{h}$ X $c^{ch}c$
3. $c^{ch}c$ X $c^{h}c$
4. $c^{ch}c^{h}$ X $c^{h}c$
5. $c^{ch}c$ X $c^{ch}c$
6. None of these
7. More than one answer is possible

b) If you cross a chinchilla rabbit with a himalayan rabbit, and the offspring are 1/2 chinchilla, 1/4 himalayan, and 1/4 albino, what were the genotypes of the parents?
1. $c^{ch}c$ X $c^{h}c^{h}$
2. $c^{ch}c^{h}$ X $c^{h}c$
3. $c^{ch}c$ X $c^{h}c$
4. $c^{ch}c^{h}$ X $c^{h}c$
5. $c^{ch}c$ X $c^{ch}c$
6. None of these
7. More than one answer is possible

B

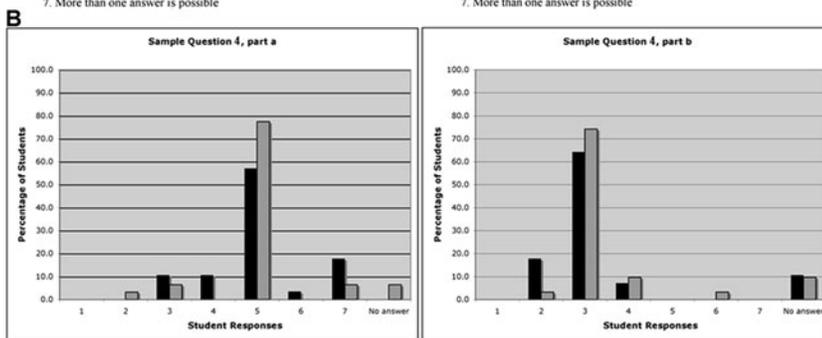**Figure 4.** Sample question 4 (A) and student responses (B). The correct response for question 4, part a, is 5; the correct response for question 4, part b, is 3. Responses from Spring 2009 are shown as black bars, and the responses from Fall 2009 are shown as gray bars. These data include only the students present in class at the time the problem was presented (28 in Spring 2009, 31 in Fall 2009).

Table 2. Exam 2 format for four consecutive semesters

Semester	Point distribution by question category[a,b]		Points per chapter[b]			Weighted mean Bloom score (±SD)
	PS/CT	FR	Extensions of Mendel	Linkage	Chromosome mutations	
Spring 2008	65	35	33	33	34	2.89 ± 0.04
Fall 2008	63	37	46	32	22	2.68 ± 0.09
Spring 2009	87	13	40	30	30	3.20 ± 0.01
Fall 2009	83	17	38	34	28	3.33 ± 0.01

[a]Questions were categorized as PS (problem solving) and CT (critical thinking) or FR (factual recall).
[b]The total point value of the exam was 100 points.

was based on class discussion. Students were informed that they would receive credit for simply answering questions (percent of questions answered) and would not be graded based on correct or incorrect answers. Thus, without the pressure to be correct, the responses are more likely to be accurate reflections of each student's understanding (James, 2006).

Exam Grades

The grades on the second exam that were analyzed in this study were the grades prior to any curving or bonus points, so the grades are true reflections of student performance. The same general exam format was used each semester (multiple-choice, fill-ins, short answer questions, and problems), but the number of each type of question and specific questions varied each semester. The second exam covered the same three chapters during all four semesters of this study, and the distribution of points for each chapter was similar in all four semesters (Table 2).

The exams used in each semester were blind-assessed by three outside evaluators in order to determine comparability of the cognitive levels of the exams. The outside evaluators are not affiliated with this genetics course but have experience teaching genetics and/or in using Bloom's taxonomy to evaluate biology questions. The evaluators did not know which exam corresponded to each semester and were not familiar with the goals of this study or any of the other data gathered. Two of the evaluators rated each question for cognitive level on a scale of 1–6 using Bloom's taxonomy (Bloom *et al.*, 1956; Crowe *et al.*, 2008). For each exam, the weighted mean Bloom score was determined as previously described (Momsen *et al.*, 2010). Briefly, the average Bloom score for each question was multiplied by the point value of the question, and the sum of these values was then divided by the total point value of the exam (Table 2). The weighted mean Bloom score was comparable between the spring and fall semesters in each calendar year, although it is clear that the 2009 exams had a higher weighted mean Bloom score than the 2008 exams. This suggests that the 2009 exams required a higher cognitive level than the 2008 exams. Two of the evaluators were also asked to categorize the questions more generally as problem solving/critical thinking or factual recall (Table 2). The 2009 exams had a significantly higher proportion of problem-solving/critical-thinking questions than the 2008 exams, again suggesting a higher cognitive level.

Attendance

Percent attendance was calculated based on the number of lectures each student attended out of the eight lectures covering material on the second exam. In the two semesters when clickers were used (Spring 2009 and Fall 2009), attendance was taken using the clicker reports. A student was assumed to be absent if he or she did not answer any questions during class. Because the classes were relatively small in size (~30), it was possible to recall whether a student who did not answer any clicker questions was actually present in class, and such students were marked present. The small class sizes also enabled monitoring whether a student was using another student's clicker to give the false impression that an absent student was present. In the two semesters when clickers were not used (Spring 2008 and Fall 2008), attendance was taken using a daily sign-in sheet. Students were not graded on attendance or penalized for absences, so there was no incentive to sign in another student who was absent, and the small class size made it possible to monitor attendance.

Correlation Analysis

Correlation analysis was used to determine whether there was a direct linear relationship between two variables, such that the variation in one data set corresponds to the variation in the other. The following pairs of data sets were compared using correlation analysis: exam grades versus percent clicker participation for each semester (Spring 2009 and Fall 2009), exam grades versus percent correct clicker responses for each semester (Spring 2009 and Fall 2009), and exam grades versus percent attendance for each semester (Spring 2008/Fall 2008 and Spring 2009/Fall 2009). Correlation coefficients were calculated using PASW Statistics, version 18. The correlation coefficient was considered to be statistically significant if $p < 0.05$.

Institutional Review

This study has been approved by the Human Subjects Committee at the University of Hartford according to conditions set forth in Federal Regulation 45 CFR 46.101(b) and was determined to be exempt from further committee review.

RESULTS

Problem Solving during Lectures

In the genetics course, the majority of points on exams are distributed among questions that require application of

problem-solving and critical-thinking skills (Table 2). Although clickers were used throughout the semester, this study focuses on exam 2 and the clicker questions utilized during the eight lectures in which the material from exam 2 was covered. The total number of clicker questions used during those eight lectures was 48 in Spring 2009 and 34 in Fall 2009.

The general approach I used to facilitate problem solving during lectures consisted of first lecturing on a concept, after which I presented the class with a novel problem that required application of that concept without discussion or explanation. After students voted for the correct answer to the problem, the histogram of student responses was displayed, and students were asked to explain which answer was correct and the rationale behind their answers. This usually consisted of a whole-class discussion, although in some instances small-group peer discussions were utilized. After the discussion period, I presented the students with a similar problem, and students again voted for the correct answer. In some instances, a third similar problem was presented, particularly if it seemed that some students were still struggling with the concept. In this article, I have included four sets of problems that were presented using this approach, but these problems and the outcomes are fair reflections of other sets of problems used throughout the semester.

In some cases, the problem and voting slide were displayed simultaneously. An example of this is shown in sample question 1 (Figure 1), which addresses the concept of codominance and multiple alleles in blood types. After a brief explanation of the genetics of blood types, students were presented with the first problem (Figure 1A, left) and were allowed to begin voting immediately (Figure 1B, left). In the Spring 2009 semester, 48% of the students answered the first question correctly; in the Fall 2009 semester, fewer than 36% of students answered correctly. After discussion of the correct answer, the second problem was displayed (Figure 1A, right) and students again voted for the correct answer (Figure 1B, right). In the Spring 2009 semester, the correct responses increased to 64%; in the Fall 2009 semester, the correct responses increased to 71%.

In other instances, I presented the problem before displaying the voting slide. In these cases, students were allowed to work for a few minutes to solve the problem without seeing the multiple-choice answers. An example of this is shown in sample question 2 (Figure 2), which involves the concept of recessive epistasis. After a description of recessive epistasis, students were shown the problem (Figure 2A), and were given sufficient time to read the problem and work on answers to parts a, b, and c. The voting slide was then displayed for part a (Figure 2B, left), and students were given time to vote for the correct answer (Figure 2C, left). In Spring 2009, 46% of the class answered the question correctly; in Fall 2009, 50% of the class answered correctly. After discussion of the correct answer, the voting slide for part b was displayed (Figure 2B, center) and students voted for the correct answer (Figure 2C, center). In Spring 2009, there was a slight increase to 57% in the correct answers, while in Fall 2009, there was a substantial increase to 82%. Finally, the answer to part b was discussed, the voting slide for part c was displayed (Figure 2B, right), and students voted for the correct answer (Figure 2C, right). In Spring 2009, there was an increase to 75% correct, while in Fall 2009, there was a slight increase to 89% correct.

The students' newfound ability to solve problems involving recessive epistasis carried over to a new problem involving the concept of dominant epistasis (Figure 3). Sample question 3 was presented in a manner similar to that of sample question 2 (Figure 2), where the students were given the problem (Figure 3A) before seeing the voting slides (Figure 3B). In Spring 2009, 68% of the class answered the first problem correctly and 79% answered the second problem correctly (Figure 3C). In Fall 2009, 86% of the class answered the first problem correctly and 82% answered the second problem correctly (Figure 3C).

Results similar to those presented in Figures 1–3 were observed with other groups of problems presented in this manner, with an increase in the percentage of correct responses on the second and subsequent problems in a series of related problems (unpublished data). However, in some instances, there was little or no appreciable increase, or even an occasional decrease, in the percentage of correct answers on the second problem. An example of this is shown in sample question 4 (Figure 4), which addresses the concept of multiple allelic series. These questions were presented immediately following an explanation of the genetic determination of coat color in rabbits. In this case, there was a slight increase from 57% to 64% correct between the first and second question in the Spring 2009 semester (Figure 4B, black bars), but a slight decrease in the percentage correct from 77% to 74% in the Fall 2009 semester (Figure 4B, gray bars).

Exam Performance

To determine whether the problem-solving skills acquired using clickers during class were reflected in student performance on exams, I analyzed exam 2 grades as a function of student clicker participation. There was a general trend of decreased exam 2 grades with decreased percent participation (Figure 5). There was only a slight decrease in average percent participation for each grade category for the "A," "B," and "C" grade ranges, although the percent participation for
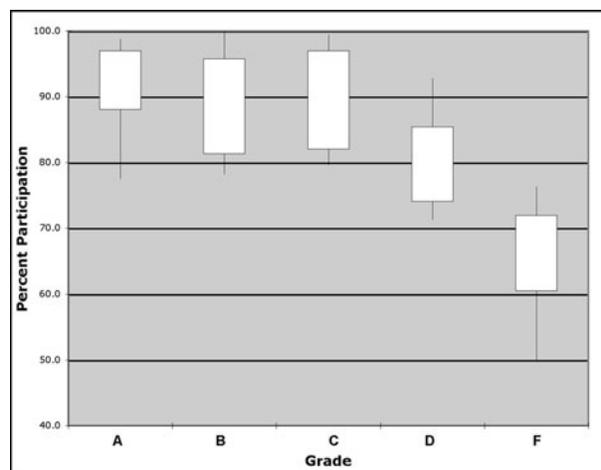

Figure 5. Student percent participation on clicker questions as a function of exam 2 grades. Data are compiled from two semesters, Spring and Fall 2009 ($n = 63$). The white bars represent the distribution of student participation for the 75th (upper limit) and 25th (lower limit) percentiles. The vertical black lines represent the 90th (upper limit) and 10th (lower limit) percentiles.

Table 3. Correlation analysis for exam 2 grade vs. clicker responses

Semester	N	Class average: exam 2	Class average: % participation[a,b]	Class average: % correct[a,c]	Correlation: exam grade vs. % participation[d]	Correlation: exam grade vs. % correct[d]
Spring 2009	30	75.3	86.3	67.7	**r = 0.518 p = 0.003**	r = 0.342 p = 0.064
Fall 2009	33	77.8	86.1	73.9	**r = 0.583 p < 0.001**	**r = 0.390 p = 0.025**

[a]These values are based on clicker questions asked in the eight lectures corresponding to material covered on exam 2.

[b]Percent participation refers to the percentage of questions each student answered out of the total questions asked (48 total questions in Spring 2009 and 34 questions in Fall 2009).

[c]Percent correct refers to the percentage of correct responses out of the total number of questions that each student answered.

[d]Values reflecting statistically significant positive correlation ($p < 0.05$) are shown in bold.

Table 4. Correlation analysis for exam 2 grade vs. attendance

Semester	N	Class average: exam 2	Class average: % attendance[a]	Correlation: exam grade vs. % attendance[b]
Spring 2008	36	77.6	94.4	r = 0.251 p = 0.140
Fall 2008	39	72.5	96.1	r = 0.179 p = 0.277
Spring 2009	30	75.3	92.9	**r = 0.567 p = 0.001**
Fall 2009	33	77.8	95.5	**r = 0.575 p < 0.001**

[a]This value is based on student attendance in the eight lectures covering material on exam 2.

[b]Values reflecting statistically significant positive correlation ($p < 0.05$) are shown in bold.

students in the "A" range displayed a smaller distribution than did the percent participation for students in the "B" and "C" range (Figure 5, white bars). However, there is a more substantial decrease in average percent participation for students in the "D" and "F" grade ranges (Figure 5).

I performed correlation analysis on the two data sets (exam grades and percent-participation grades) in order to determine the degree to which the variation in exam grades reflected the variation in clicker participation. In both semesters, there was a statistically significant positive correlation between the exam grades and percent-participation grades (Table 3). To determine whether the percentage of clicker questions a student answered correctly correlated with their exam grades, I also performed correlation analysis between these two data sets. There was not a statistically significant positive correlation between exam grades and percent correct clicker responses in Spring 2009, but in Fall 2009, there was a statistically significant positive correlation (Table 3).

Attendance and Clicker Participation

The percent-participation grades were calculated based on the number of questions that a student answered out of the total number of clicker questions asked during the eight lectures in which exam 2 material was covered. If a student did not answer a question, it was because either the student was in class and did not answer in time or the student was not in class at the time the question was asked. Thus, it is possible that those students with low participation grades performed poorly on the exam because of poor attendance alone. To address this, I determined a percent-attendance grade for each student based on the number of lectures each student attended out of the eight lectures in which exam 2 material was covered, and performed correlation analysis between the

exam 2 grades and percent attendance. In Spring 2009 and Fall 2009, there was a statistically significant positive correlation between exam grades and attendance (Table 4), suggesting that a poor performance on the exam may have been due to poor attendance rather than poor clicker participation. However, when I performed correlation analysis between exam grades and percent attendance for the two semesters when clickers were not used (Spring 2008 and Fall 2008), the correlation coefficients were not statistically significant (Table 4).

The discrepancy between attendance and exam performance in the presence and absence of clickers can best be explained by analyzing the number of students who attended all eight lectures in which exam 2 material was covered, yet failed the exam. There was not a significant change in the percentage of students who failed the exam in the absence and presence of clickers (13.3% in Spring and Fall of 2008 vs. 11.1% in Spring and Fall of 2009). In the two semesters when clickers were not used, 10 students failed exam 2, but seven of them were present for every lecture (Table 5). However, in the two semesters when clickers were used, only one out of the seven students who failed was present at every lecture.

Table 5. Failing exam 2 grades for students with perfect attendance

Semesters	"F" with 100% Attendance[a]	Total "F"
Spring and Fall 2008	7	10
Spring and Fall 2009	1	7

[a]Based on student attendance in the eight lectures covering material on exam 2.

DISCUSSION

Using Clickers to Teach Problem-Solving Skills

Before using clickers, my approach to teaching problem solving was to present students with sample problems to solve during class. After allowing time for the students to solve the problems (sometimes independently and sometimes in small groups), I asked the students to explain how they solved the problem. I frequently observed that the same students were always the first to volunteer to share their answers with the class, while other students were reluctant to do so, even when they were directly called upon. I also observed that some students would not even attempt to solve the problem and would instead wait for other students to explain the solution. This was reflected in some of the clicker evaluation comments written by students (see *Student Feedback* section below). A potential complication generated by this apparent apathy is that students who do not attempt sample problems in class are less likely to develop the problem-solving skills necessary for success on exams.

I decided to use clickers in order to promote independent problem solving, while still encouraging discussion to reinforce concepts. My approach was to explain a concept and then present the students with a novel problem requiring application of that concept without explanation or discussion of how to solve the problem. Encouraging students to attempt the problem independently before discussing it enabled them to solve the problem without the bias that could be generated by stronger students giving convincing arguments to persuade weaker students toward the correct (or sometimes incorrect) answer. If a student initially chooses the incorrect answer, they can learn from their mistakes and identify their own weaknesses. After observing the histogram of student responses and discussing the correct answer, weaker students may experience increased confidence after they are able to understand and solve something that a few minutes earlier seemed insurmountable. This is reflected in the general increase in the number of students answering second and subsequent problems in a set correctly (Figures 1–3).

There was semester-to-semester variation in the extent to which students learned how to solve a certain type of problem. For example, in sample question 2 (Figure 2), students in Fall 2009 seemed to understand how to solve the problem by the second attempt, as indicated by an increase from 50% to 82% correct responses (Figure 2C, left and center, gray bars). However, in Spring 2009, there was only a slight increase on the second attempt from 46% to 57% (Figure 2C, left and center, black bars). By the third attempt, 89% of the students answered the question correctly in Fall 2009 (Figure 2C, right, gray bars) and 75% of the students answered the question correctly in Spring 2009 (Figure 2C, right, black bars). Thus, despite the fact that students in both semesters performed similarly on the first question, the Fall 2009 students achieved a much more rapid understanding of the problem. When a set of problems on a related concept was presented (Figure 3), the percentage of students answering the first question correctly was similar to the percentage of students answering the third question from the previous set of problems correctly (Figure 2). Students retained their understanding of how to solve this type of problem and were able to answer a similar but different problem correctly on the first attempt.

There was not always an increase in the percentage of students answering the follow-up questions correctly (Figure 4). This is not surprising, as in many cases the second question was more challenging than the first question. Gradually increasing the difficulty level of the questions is important for promoting development of problem-solving and critical-thinking skills, but this increase will be reflected in an occasional decrease in the number of students answering the follow-up questions correctly. However, there was often an increase in the percentage of students answering correctly, even when the second problem was slightly more difficult than the first (Figure 2).

I chose to show the histogram of student responses, but not the correct answer, prior to the discussion period. A recent study showed that seeing the histogram of student responses can weaken and bias discussions because most students will assume that the most common answer is the correct answer (Perez *et al.*, 2010). In that particular study, students were allowed to vote on a clicker question, discuss the question, and revote on the same question. When students saw the histogram, they were more likely to change their answer to the most popular answer. While it is true that seeing the histogram from the first problem may have affected the quality of discussion in my study as well, the fact that the follow-up question was similar to, but different from, the original question meant that students needed to understand how to solve the problem in order to choose the correct answer on the follow-up question.

Previous studies have shown that small-group peer instruction, in which students work together in small groups to solve a problem, is an effective method for encouraging involvement of all students in problem-solving activities (Knight and Wood, 2005; Crouch *et al.*, 2007; Preszler, 2009). It has been shown that peer instruction is more effective than classwide discussion in promoting conceptual understanding in large classes (Nichol and Boyle, 2003). Additionally, the combination of peer instruction and instructor explanation has been shown to be more effective than either peer instruction or instructor explanation alone (Smith *et al.*, 2011). I chose to use both small-group discussion and classwide discussion throughout this study simply because the small size of my classes made classwide discussion feasible. However, it may have been even more effective to encourage students to discuss the answer to the first problem in small groups in all instances, and I will likely incorporate such an approach more frequently in the future. In a similar study, it was observed that peer discussion in small groups after answering a clicker question increased the ability of students to answer a second similar question correctly (Smith *et al.*, 2008). This result was observed even when none of the students in a group answered the first question correctly. While that study offers strong support to the assertion that small-group peer discussion does improve student understanding of concepts, that study was completed in a class of 350 students, so the classwide approach might not have been very effective.

Clicker Participation Correlates with Exam Performance

The results of this study demonstrate that student engagement in problem solving and critical thinking during lectures, as measured by percent clicker participation, correlates with

student performance on an exam stressing problem-solving and critical-thinking skills (Table 3 and Figure 5). Interestingly, there was no clear correlation between exam grades and the percentage of clicker questions that a student answered correctly (Table 3). This suggests that the value of this approach lies in students simply attempting to solve the problems, not in answering the questions correctly during class.

Clicker participation was factored into the students' semester grades through the use of a class-participation grade. This grade constituted 5% of the total semester grade, and was based not only on the percentage of clicker questions answered, but also on the extent to which each student was involved in class discussions through asking and answering questions. So that students would not feel an incentive to cheat on clicker questions, I informed them that they would not be graded on correct versus incorrect responses, but rather on the percentage of clicker questions that they answered. Previous studies have shown that such a low-stakes grading system minimizes the number of students relying on others to answer questions, and thus student responses are a pure reflection of each student's understanding (James, 2006). In fact, I observed that even when I told the students that they could discuss a particular problem with their neighbors before voting, most students preferred to attempt the problem on their own.

The best explanation for this correlation between percent-participation and exam grades comes from observing the fact that students in the "D" and "F" grade ranges had lower percent participation than students in the "A," "B," and "C" ranges (Figure 5). It is noteworthy that the average percent participation did not vary significantly among the "A," "B," and "C" students. This is not surprising, because any student present in class could easily answer all of the questions, even if he or she was blindly guessing or did not truly understand the concepts. The variation in grades among students with high percent participation may represent an overall difference in academic abilities, a difference in their efforts in studying outside of class, and/or a difference in their level of engagement during class. It is clear that students in the "A" range have a more narrow distribution of percent participation than students in the "B" and "C" ranges (Figure 5), which may indicate that the "A" students had the highest level of overall engagement during class. Similar trends were reported in a previous study (Perez et al., 2010), although the variation in participation levels at each grade range was more pronounced in that study. This difference is likely due to a significant difference in the number of students involved in the two studies (629 compared with 63 in this study), as well as a difference in the total number of clicker questions analyzed.

The percent clicker participation grade was calculated based on the number of questions that a student answered out of the total number of questions asked. If a student had a low percent-participation grade, it could have been because he or she was not present in class, rather than because of a lack of engagement. When percent attendance was compared with exam grades, there was a statistically significant correlation between the two values (Table 4), indicating that the decrease in student performance may have been due more to the lack of attendance than to the lack of involvement in answering clicker questions. When the percent attendance was compared with exam grades in two semesters when clickers were

not used (Spring and Fall 2008), there was not a statistically significant correlation (Table 4). This observation suggests that the use of clickers increased the extent to which students present in class were able to develop the problem-solving skills necessary for success on exams. Consistent with this, in the two semesters when clickers were not used, 70% of the students who failed the exam had been present at every lecture in which the material on that exam was covered, whereas in the two semesters when clickers were used, only 14% of students failed, despite having been present at every lecture (Table 5). This suggests that, prior to the use of clickers, those students who did not understand the concepts did not attempt to solve the sample problems given during class, which carried over into an inability to solve problems on the exam. It is likely that there were students in previous semesters who were present in class but were simply not engaged during lectures. This could have stemmed from general apathy or disinterest in the subject or from frustration and a sense of being overwhelmed by material they did not understand. It appears that using clickers promoted increased engagement of those students who would otherwise not have participated in problem-solving activities in class.

One possible explanation for the difference in the relationship between exam performance and attendance between 2008 and 2009 is that the 2009 semesters may have consisted of students with higher academic abilities than the 2008 semesters. This is unlikely, because the GPAs of students entering the class in all four semesters were not significantly different (Table 1). In fact, one of the semesters when clickers were not used (Fall 2008) had the highest class GPA of the four semesters in the study, and one of the semesters when clickers were used (Spring 2009) had the lowest (Table 1).

It is interesting that there was not a significant difference in the percentage of students who failed the exam from the semesters when clickers were not used (13.3% in Spring and Fall 2008) to the semesters when clickers were used (11.1% in Spring and Fall 2009). It is not the percentage of students failing that decreased, just the percentage of students who had been present in class and still failed. The students failing in Spring and Fall 2009 were almost exclusively students with poor attendance (Table 5). It is also worth noting that the use of clickers had no effect on the class averages on the exams (Table 4). If the use of clickers enhanced student understanding of material, one would think that this would be accompanied by an overall increase in class averages on the exam and a decrease in the percentage of students failing the exam. However, the exams in Spring and Fall 2009 required a higher cognitive level compared with the previous two semesters (Table 2), so this likely offset the improved ability to solve problems in the determination of exam grades. Considering that the exams were more difficult in the semesters when clickers were used than in the semesters when clickers were not used, it is all the more striking that there was a positive correlation between exam performance and attendance in the 2009 semesters and not in the 2008 semesters.

Homework

Success in a course stressing problem solving and critical thinking requires that students not only are involved in problem solving during lectures but that they also practice solving problems outside of class to ensure that they have the ability

to think critically and independently to apply concepts in different contexts. In all four of the semesters addressed in this study, homework problems were assigned. These included short problem sets consisting of approximately two to three questions for each chapter, as well as textbook practice problems consisting of 10–20 questions for each chapter. Of these, only the short problem sets were collected and graded.

One possible explanation for the correlation between clicker participation and exam grades is that the involvement of students in solving problems during class increased the likelihood that they would attempt to complete practice problems outside of class. In this scenario, by being involved in problem solving in class, a student would gain the confidence to attempt problems independently outside of class, which would be reflected in exam performance. In all four semesters, almost all students completed and submitted the short problem sets, because this was a graded assignment (unpublished data). There was no significant trend of grade improvement on these homework assignments in the semesters where clickers were used compared with the semesters when clickers were not used (unpublished data). Because students are allowed to work together on the graded problem sets, the grades do not necessarily reflect each individual student's effort level and understanding. Additionally, the short length of these problem sets means that if these are the only problems that a student completes, he or she will not have ample opportunity to fully develop problem-solving abilities.

The textbook practice problems serve the important function of exposing students to a wide array of different types of problems. I did not collect these problems, because the answers are available in the back of the book, but students were informed that they were expected to attempt these problems. Because these problems were not graded, the incentive to complete them came solely from their value as preparation for exams. It is possible that the use of clickers stimulated an increase in the number of students independently attempting these practice problems outside of class. However, information regarding the number of students completing the practice problems was not gathered during the four semesters in this study.

Student Feedback

To reach the conclusion that the use of clickers enhanced student engagement and learning in the classroom, the opinions of the students must be taken into consideration. The students completed clicker evaluations at the end of each semester (Fall and Spring 2009), and feedback was overwhelmingly positive. When asked to rate the statement "overall, I feel that the use of clickers enhanced my learning experience in this class" on a scale of 1 (strongly disagree) to 5 (strongly agree), the average rating was 4.3. Students were asked to explain what they liked most about the use of clickers, and the following quotations represent typical responses:

> "It allows me time to figure out the answer on my own rather than relying on someone else."

> "Without a clicker, I might not try so hard to solve the problem. I might be lazy and wait for an answer. Actually getting the time to think and use the clicker helped a lot."

> "It allows students to do sample problems while the lecture is still fresh in their minds."

> "It keeps me interested instead of just listening to lecture, gives me an opportunity to see if I really understand."

> "I don't have to answer out loud so I don't feel completely dumb if I am wrong."

> "They keep the class more relaxed. We are able to discuss the answers to the questions and how each of us came up with the answer that we did."

However, for some of the top students who were able to solve the initial problems correctly, the repetition as well as the time it took other students to respond was a source of some frustration. When asked what they liked least about the use of clickers, most students did not respond. However, a few students responded with comments similar to the following:

> "Takes too long to get results because people take time to answer questions."

> "Sometimes I feel like I don't have enough time."

> "Some people may need more time than others to complete the problems."

Thus, while some students felt that too much time was spent on each clicker problem, other students felt that they did not have enough time. To address these problems, I have tried to find the balance between allowing enough time for all students to answer without making the top students wait for too long. To accomplish this, I initiated a countdown timer once about 75% of the students had entered their responses and limited the number of clicker questions used in each lecture.

CONCLUSIONS

The use of clickers gives students the opportunity to attempt to solve problems independently during lectures. This promotes development of problem-solving skills, particularly in those students who would normally be too intimidated to attempt the problem in a class or group discussion forum. This is reflected in the correlation between percent clicker participation and exam grades. Importantly, this does not matter whether students answer questions correctly during class, just that they are attempting to answer questions. By engaging in problem solving during class, students will gain confidence in their problem-solving abilities and will thus be more likely to be able to solve problems on exams. One possible explanation for this is that the increased confidence makes the students more likely to attempt to solve textbook practice problems outside of class, which in turn improves exam performance. This may be an avenue for further investigation.

ACKNOWLEDGMENTS

statistical analyses and to A. Crowe, C. Paul, and K. MacLea for evaluating exams.

REFERENCES

Allen D, Tanner K (2002). Approaches in cell biology teaching. Cell Biol Educ *1*, 3–5.

Armstrong N, Chang SM, Brickman M (2007). Cooperative learning in industrial-sized biology classes. CBE Life Sci Educ *6*, 163–171.

Bloom BS, Krathwohl DR, Masia BB (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain, New York: Longmans, Green.

Cain J, Robinson E (2008). A primer on audience response systems: current applications and future considerations. Am J Pharm Educ *72*, 77.

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

Chi MTH, De Leeuw N, Chiu M, Lavancher C (1994). Eliciting self-explanations improves understanding. Cogn Sci *18*, 439–477.

Coleman EB (1998). Using explanatory knowledge during collaborative problem solving in science. J Learn Sci *7*, 387–427.

Coleman EB, Brown AL, Rivkin ID (1997). The effect of instructional explanations on learning from scientific texts. J Learn Sci *6*, 347–365.

Collins J (2008). Audience response systems: technology to engage learners. J Am Coll Radiol *5*, 993–1000.

Collins LJ (2007). Livening up the classroom: using audience response systems to promote active learning. Med Ref Serv Q *26*, 81–88.

Crossgrove K, Curran KL (2008). Using clickers in nonmajors- and majors-level biology courses: student opinion, learning, and long-term retention of course material. CBE Life Sci Educ *7*, 146–154.

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

Crouch CH, Watkins J, Fagen AP, Mazur E (2007). Peer instruction: engaging students one-on-one, all at once. In: Reviews in Physics Education Research, ed. EF Redish and P Cooney, College Park, MD: American Association of Physics Teachers, 1–55.

Crowe A, Dirks C, Wenderoth MP (2008). Biology in bloom: implementing Bloom's taxonomy to enhance student learning in biology. CBE Life Sci Educ *7*, 368–381.

DeBourgh GA (2008). Use of classroom "clickers" to promote acquisition of advanced reasoning skills. Nurse Educ Pract *8*, 76–87.

Draper SW, Brown MK (2004). Increasing interactivity in lectures using an electronic voting system. J Comp Assist Learn *20*, 81–94.

Herreid CF (2010). It's all their fault. J Microbiol & Biol Educ *11*, 34–36.

James MC (2006). The effect of grading incentive on student discourse in peer instruction. Am J Phys *74*, 689–691.

Knight JK, Wood WB (2005). Teaching more by lecturing less. Cell Biol Educ *4*, 298–310.

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

Momsen JL, Long TM, Wyse SA, Ebert-May D (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. CBE Life Sci Educ *9*, 435–440.

Nichol DJ, Boyle JT (2003). Peer instruction versus classwide discussion in large classes: a comparison of two interaction methods in the wired classroom. Stud High Educ *28*, 457–473.

Perez KE, Strauss EA, Downey N, Galbraith A, Jeanne R, Cooper S (2010). Does displaying the class results affect student discussion during peer instruction? CBE Life Sci Educ *9*, 133–140.

Preszler RW (2009). Replacing lecture with peer-led workshops improves student learning. CBE Life Sci Educ *8*, 182–192.

Preszler RW, Dawe A, Shuster CB, 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.

Reay NW, Li P, Bao L (2008). Testing a new voting machine question methodology. Am J Phys *76*, 171–178.

Russell JS, McWilliams M, Chasen L, Farley J (2011). Using clickers for clinical reasoning and problem solving. Nurse Educ *36*, 13–15.

Smith MK, Wood WB, Adams WK, Wieman C, Knight JK, Guild N, Su TT (2009). Why peer discussion improves student performance on in-class concept questions. Science *323*, 122–124.

Smith MK, Wood WB, Knight JK (2008). The Genetics Concept Assessment: a new concept inventory for gauging student understanding of genetics. CBE Life Sci Educ *7*, 422–430.

Smith MK, Wood WB, Krauter K, Knight JK (2011). Combining peer discussion with instructor explanation increases student learning from in-class concept questions. CBE Life Sci Educ *10*, 55–63.

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

Tanner K, Chatman LS, Allen D (2003). Approaches to cell biology teaching: cooperative learning in the science classroom—beyond students working in groups. Cell Biol Educ *2*, 1–5.

Tanner KD (2009). Talking to learn: why biology students should be talking in classrooms and how to make it happen. CBE Life Sci Educ *8*, 89–94.