

Article

Does Displaying the Class Results Affect Student Discussion during Peer Instruction?

Kathryn E. Perez,* Eric A. Strauss,* Nicholas Downey,* Anne Galbraith,* Robert Jeanne,[†] and Scott Cooper*

*Department of Biology, University of Wisconsin at La Crosse, La Crosse, WI 54601; and [†]University of Wisconsin–Madison, Madison, WI 53706

Submitted November 13, 2009; Revised January 21, 2010; Accepted February 6, 2010
Monitoring Editor: Mary Pat Wenderoth

The use of personal response systems, or clickers, is increasingly common in college classrooms. Although clickers can increase student engagement and discussion, their benefits also can be overstated. A common practice is to ask the class a question, display the responses, allow the students to discuss the question, and then collect the responses a second time. In an introductory biology course, we asked whether showing students the class responses to a question biased their second response. Some sections of the course displayed a bar graph of the student responses and others served as a control group in which discussion occurred without seeing the most common answer chosen by the class. If students saw the bar graph, they were 30% more likely to switch from a less common to the most common response. This trend was more pronounced in true/false questions (38%) than multiple-choice questions (28%). These results suggest that observing the most common response can bias a student's second vote on a question and may be misinterpreted as an increase in performance due to student discussion alone.

INTRODUCTION

Peer instruction involves giving students a problem to solve as a group, with much of the teaching occurring among peers. Peer instruction has been demonstrated to result in higher learning gains than lecture in many different courses (MacManaway, 1970; Hake, 1998; Rao and DiCarlo, 2000; Pollock, 2006), including introductory biology (Knight and Wood, 2005; Armstrong *et al.*, 2007; Freeman *et al.*, 2007; Preszler *et al.*, 2007; Crossgrove and Curran, 2008; Walker *et al.*, 2008; Armbruster *et al.*, 2009). Two challenges with peer instruction in large lecture courses include monitoring whether the students are actually discussing the problem and then collecting their responses. The use of personal response systems (commonly known as clickers) coupled with peer instruction has helped alleviate these problems. Pioneered for physics courses in large part by Mazur (1997), the technique has since spread to most disciplines, combining the benefits of peer instruction and formative feedback

in a way that is more effective than lecture (Hake, 1998; Preszler *et al.*, 2007).

The use of peer instruction and clickers to provide formative assessment has several advantages over lecture alone. A positive correlation between the number of in-class questions given and overall course performance has been reported for multiple biology courses, suggesting that frequent formative assessment can enhance learning (Preszler, 2009). A controlled experiment comparing formative assessment alone with formative assessment and peer discussion revealed that peer discussion adds further to the value of in-class problem solving (Smith *et al.*, 2009). Other studies show that understanding is more likely to develop when students engage in activities such as analysis, evaluation, interpretation, prediction, and explanation (Coleman *et al.*, 1997; Coleman, 1998; Bransford *et al.*, 1999).

Some analyses have shown that students giving the explanations in a peer group show greater learning gains than those receiving the explanation (Webb, 1989; Coleman *et al.*, 1997), suggesting that the active process of explaining forces a student to integrate new knowledge with existing knowledge (Chi *et al.*, 1994). In addition to benefits to the students doing the explaining, the recipients may also benefit from peer instruction because their peers come from a similar

DOI: 10.1187/cbe.09-11-0080

Address correspondence to: Kathryn E. Perez (perez.kath@uwlax.edu).

background and may be better at clearing up misconceptions or finding relevant examples than their instructor (Wood, 2004). Smith *et al.* (2009) tested whether learning gains from student discussion were due to the process of constructing knowledge as part of the discussion, or due to peer influence by students in the group who seemed knowledgeable about the correct answer. They found that participation in a discussion group alone led to learning gains even if no one in the group originally knew the correct answer. This indicates that the discussion process and construction of individual student knowledge leads to learning gains for all members of the group.

One common classroom model for engaging students in peer instruction is to ask them to use their clickers to independently answer a formative assessment question, either true/false (two-choice) or multiple choice (four-choice), show them the class response (but not the correct answer), and have students work in groups to reach a consensus and then vote a second time. This think-pair-share (TPS) approach (Lyman, 1981; Allen and Tanner, 2002) is designed to enable students to arrive at a better understanding of the question by discussing it with their neighbors (Mazur, 1997; Crouch and Mazur, 2001; Slater *et al.*, 2006). The TPS approach has been shown to be effective in improving student in-class performance (Nichol and Boyle, 2003; Smith *et al.*, 2009). The benefits of peer instruction also extend beyond the classroom: students perform better when assessed on both content knowledge and on their problem-solving skills (Crouch *et al.*, 2007).

In one variation of the TPS approach, the bar graph of initial student responses is shown before student discussion. “The peer learning model (also known in the literature as peer instruction or PI), requires that students think and answer independently first, see the answers, and then spend time in groups struggling to reach a consensus answer” (Caldwell, 2007). Although display of the histogram is not part of their study, Smith *et al.*, 2009 describe the technique as follows: “When PI is used, students are first asked to answer a question individually, and then a histogram of their responses may be displayed to the class” (Smith *et al.*, 2009). Other authors suggest that projecting the bar graph of class responses allows students to see where their personal answer falls in the group (Preszler *et al.*, 2007), and surveyed students have indicated they like the reassurance that they are not alone in their thinking even when they are wrong (Beatty, 2004). Beatty (2004) includes bar diagram display as a component step in his description of the Question Cycle—an effective model for class communication system use (Beatty, 2004).

In a classical study in social psychology, Asch (1951) demonstrated that some students adopt the majority response, even if they know it to be incorrect. A large body of subsequent work has found consistent support for the phenomenon of group conformity (for review, see Bond and Smith, 1996). In the context of a classroom peer discussion, the bar graph of student responses could represent strong whole-course peer influences and because of an unconscious desire to conform, students might not consider all answers equally, thus lowering the quality of peer discussion. Our study asks: are students choosing their answer after peer discussion because the discussion increases their under-

standing, or are they simply conforming to the most common answer shown in the initial class-response graph?

METHODS

Course Background

General Biology is a 100-level course taken by biology majors at the University of Wisconsin at La Crosse. In fall 2008, the class consisted of eight lecture sections with a maximum of 95 students each. The seven instructors met weekly and shared a unified set of teaching materials. Six of the instructors each taught one section and one instructor taught two sections of the class. The course includes sections on ecology, cell biology, genetics, and evolution. The lecture material is arranged in a series of learning cycles, with short segments of lecture interspersed with problem solving and clicker questions on basic concepts. At the end of each unit, a series of clicker case studies are used, in which students have to integrate and apply the unit’s concepts. Each of the instructors used clickers, TPS, and display of the histogram throughout the semester. However, this was not done in a coordinated or regulated manner before the study. The data used for this analysis were gathered during the final unit of the class on evolution. Data from all eight sections were combined for the analyses summarized below. Our total sample consisted of the 629 students who signed permission waivers.

The clicker used in this study was the most current model of radio frequency personal response system called iClicker (www.iclicker.com). Students received clickers as part of their textbook rental package and then registered them online. Registered clickers could then be linked to individual students in the grade database. Points were not assigned for clicker participation; however, students were informed that in case of a borderline grade, active clicker participation could be used to boost them into the higher grade category. A low-stakes approach to clicker participation was used to more accurately reflect the content knowledge of each student participant as demonstrated by James (2006) and because student conformity is increased by task importance (Baron *et al.*, 1996). A Mann–Whitney *U* test was used to compare clicker participation between treatments to examine any potential for bias due to this low-stakes approach.

For this study, we developed 18 clicker questions to be used in all eight lecture sections. Nine of these questions were multiple-choice (four possible answers) questions, and nine were true/false (two possible answers) questions. These questions were interspersed throughout the final 3-wk unit of the class on evolution along with one to nine other nonscored questions per class period lasting 55 or 85 min. For each scored question we used the TPS approach: students were allowed to vote initially and were then given an opportunity to discuss the question with their neighbors before revoting. The bar graph of student responses was not shown while the 30-s voting period was taking place (sample question and bar graph shown in Figure 1). The treatments in this study consisted of varying which sections saw the bar graph after the initial voting period but before discussion and revoting, and they were set up in a crossover design. Each question was asked in all eight sections, but in half of the sections the students were shown the bar graph

Mules are the sterile offspring of the mating of a horse with a donkey. What would you predict would be the ultimate long-term effect of sexual selection acting on these two species?

- A. Break-down of reproductive isolating mechanism and the merging of the two species into one.
- B. The evolution of more effective postzygotic isolating mechanisms.
- C. The evolution of effective prezygotic isolating mechanisms.
- D. There would be no change in sexual selection and a horse would be just as likely to mate with a donkey as with another horse.

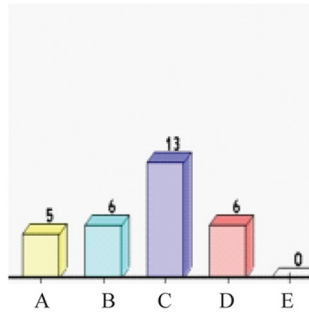


Figure 1. Study question 3.2 (left) and bar graph of student responses (right). In half of the sections students would have seen the bar graph before discussion and revoting.

before initiating discussion followed by revoting and in half they were not (Table 1). A crossover design was used in the experiment, in which student responses were shown for half of the questions and not shown for the other half in four of the lecture sections. The other four lecture sections had the reciprocal arrangement of displayed student responses. Clicker questions not used as part of the study were also interspersed throughout these lectures to identify misconceptions, break up lecture, or for concept checks.

After completion of the class, iClicker reports were generated using the iGrader 5.0.3.1 software (www.iclicker.com). These HTML reports included every clicker vote for each student, and a corresponding screen capture to identify which question was being asked. These data were exported to Excel (Microsoft, Redmond, WA) and matched to student names and final percentage points earned in the class. Student grades were based on four lecture exams, weekly quizzes, two out-of-class group assignments, and the Introductory Biology lab. Grades were assigned using the following percentage scale: “A” (90–100), “B” (89–80), “C” (79–70), and “D/F” (<69).

We began this study with a set of 18 questions. For some questions a large percentage (95%) of students answered correctly on the first vote. In those cases, when students saw the bar graph of class responses, it seemed too artificial to initiate discussion with 95% of the class having the correct answer. Therefore, some instructors did not re-ask the question. In addition, because of time constraints at the end of the semester, a few questions were not asked by all instructors. To eliminate the impact of questions that were not asked twice in a section, they were excluded from our analysis. Most questions could be retained, as they were asked in both treatments. However, two questions that were not asked a second time at all or in only one section were dropped

Table 1. Structure of the crossover design

Question	Lecture section	
	1–4	5–8
Odd-numbered questions	Bar graph shown	Bar graph not shown
Even-numbered questions	Bar graph not shown	Bar graph shown

completely. In addition, a draft version of one question that had five possible multiple-choice responses rather than our study’s standard four choices was accidentally asked in three of the eight sections and could not be compared. Removal of these three questions left 15 questions for the analysis (Figure 2).

Using those 15 questions, we compared the responses of students who saw the bar graph of class responses with those who did not. Although students in both treatments discussed the answers between responses, the students who saw the initial bar graph of class responses knew which answer was the most common but did not know whether it was also correct. We placed the responses into two categories for each clicker question based on the bar graph from the initial class response to a question: most common (MC) and less common (LC, all other responses). Students who did not respond twice to a clicker question (pre- and post-discussion) were eliminated from the analysis for that question. Because the proportion of students falling into each category varied across questions, we examined the percentage of students in each of four combination categories: MC to MC, MC to LC, LC to MC, and LC to LC. For example, LC to MC would be a student initially voting for a less common answer for the first vote and then after discussion, voting for the most common answer (Table 2).

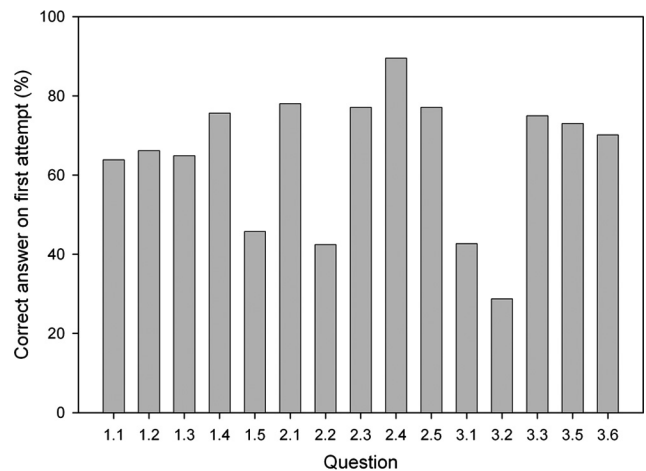


Figure 2. For each of the 15 questions, the number of students who answered correctly on the first attempt.

Table 2. Answer coding for possible student responses to first and second instance of questions

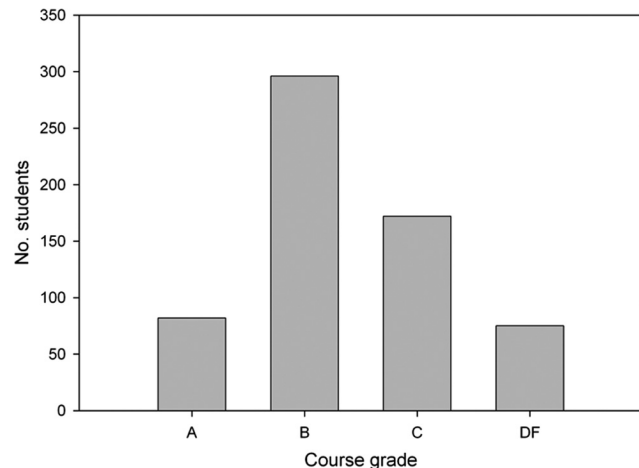
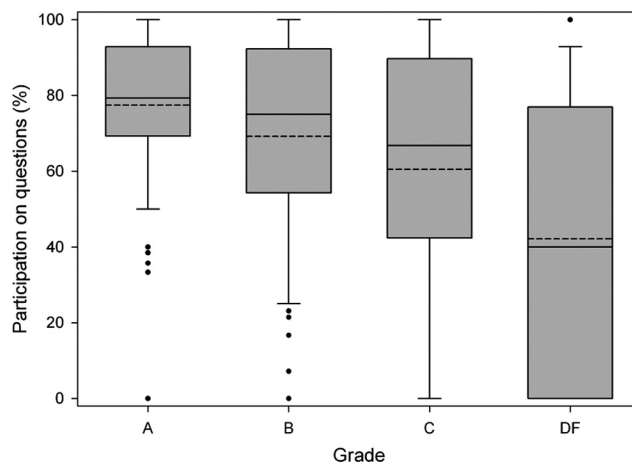
Answer coding	Student first response	Student second response
MC-MC	C	C
MC-LC	C	A, B, D
LC-MC	A, B, D	C
LC-LC	A, B, D	A, B, D

The class sections that saw the graph of class responses, as displayed in Figure 1, would see that C is the most common answer.

In total, 4182 student-response combinations were analyzed. Multiple-choice and true/false questions were examined both separately and combined, using chi-square (χ^2) tests in SAS 9.1 (SAS Institute, Cary, NC) to test the null hypothesis of no association (i.e., no effect) between the row variable (student response to question) and the column variable (question type, bar graph displayed, or student grade). When student grade was used as a variable, the Mantel-Haenszel chi-square (χ^2 M.-H.) statistic was used because "grade" is an ordinal scale variable (SAS Institute, 2004). Student engagement was quantified as the percentage of the 15 questions each student answered during the study.

RESULTS

Of the 629 students in our sample, 91 received "A's", 305 "B's," 168 "C's," and 65 "D/F's" for their final course grades (Figure 3). There was a positive relationship between student engagement, as assessed by clicker participation on the study questions, and student grade in the class (Figure 4). Students who received "A's" in the class had an average participation on clicker questions that was twice that of students who received grades of "D" or "F" (Figure 4). We have no way of knowing whether the reason for the rela-

**Figure 3.** Distribution of course letter grades. Grades were assigned using the following percentage scale: "A" (90–100), "B" (89–80), "C" (79–70), and "D/F" (<69).**Figure 4.** Participation as a function of student final course grades. The gray boxes represent the 25th and 75th percentiles. Whiskers represent the 10th and 90th percentiles. Solid lines represent the median and dashed lines represent the mean percentage participation. Dots represent outliers that fell outside the whiskers.

tively low participation of "D" and "F" students was that they were present in class and not using their clickers or that they were absent from class. There was no significant difference in clicker participation among the treatments in overall clicker participation, participation on first click, or participation on second click ($p > 0.05$ in all three cases).

For most questions, the majority of students did not change their response after peer discussion, with 56% of students (percentages are means over both treatments) choosing the most common answer both times and 21% choosing a less common response both times (Figure 5A, outer two categories). Of those who changed their response, a greater proportion of students switched to the most common answer (14%) than switched to a less common answer (9%), regardless of whether they saw the class bar graph (LC to MC vs. MC to LC, Figure 5A, middle two categories). If students saw the bar graph, they were 30% more likely to switch from LC to MC than were those who did not see the graph, a significant difference ($\chi^2 = 42.87$; $p < 0.0001$; $N = 4182$) (Figure 5A, third category). Because we were most interested in those students who changed their answer after the discussion, and the possible influence of seeing the class responses, we specifically examined data from students who changed their answer after initially answering incorrectly (LC to MC). This analysis revealed that those who saw the bar graph switched to the most common answer significantly more often than those who did not see the bar graph ($H_0: 50\%:50\%$; $\chi^2 = 42.51$; $p < 0.0001$; $N = 625$; Figure 5B). In the single question where MC was not also correct (question 1.5), more students switched from a correct answer to an incorrect but more common answer when they saw the bar graph. However, this was based on a small number of students ($N = 5$).

We next asked whether there was any difference in student response between students who saw the bar graph and those who did not in answering true/false questions with two possible answers (two-choice) or multiple-choice questions (four-choice). There was still a significant difference

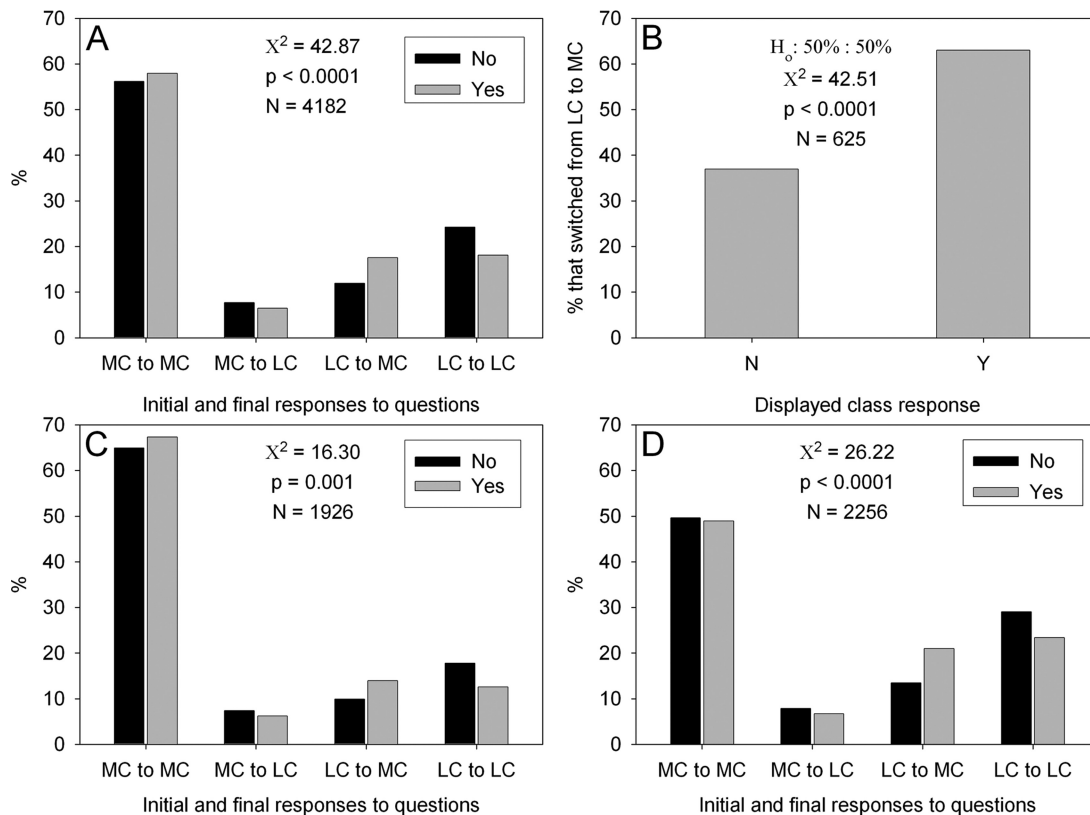


Figure 5. Comparison of student responses between those who saw the bar graph of class responses before discussion and those who did not. The bars within each panel show the percentage of students whose responses fit into each answer category. “No” indicates students did not see the bar graph before discussion and revoting (No: $N_{No} = 1933$, $N_{MCMC} = 1086$, $N_{MCLC} = 148$, $N_{LCMC} = 231$, $N_{LCLC} = 468$). “Yes” indicates students did see the bar graph (Yes: $N_{Yes} = 2249$, $N_{MCMC} = 1303$, $N_{MCLC} = 145$, $N_{LCMC} = 394$, $N_{LCLC} = 407$). A and B include combined data for both two-choice (T/F) and four-choice (multiple-choice) questions. (A) Differences in student revoting based on viewing the bar graph before discussion and revoting. (B) Percentage of students that switched from an LC to the MC answer after seeing the response bar graph (Y) or not (N). (C) Comparison of responses for true/false (two-choice) questions only. (D) Comparison of responses for multiple-choice (four-choice) questions only.

between treatments in students switching from LC to MC for both two-choice and four-choice questions (Figure 5C: $\chi^2 = 16.30$; $p = 0.001$; $N = 1926$; Figure 5D: $\chi^2 = 26.22$; $p \leq 0.0001$; $N = 2256$). The influence of seeing the bar graph was stronger for the two-choice than for the four-choice questions, with 38% more changing from LC to MC when given two choices (Figure 5C) versus a 28% increase when given four choices (Figure 5D).

We also wanted to determine whether the effect of seeing the bar graph displayed before discussion varied with final grade in the class. For this analysis, we looked only at the data for those who switched their answers. We did not find a significant effect of student grade on the percentage of students who switched from an LC to the MC answer ($\chi^2 = 1.04$; $p = 0.3087$; $N = 625$; Figure 6A). Students of all grade levels seemed to be influenced by seeing the most common answer. However, we did find a significant effect of grade on the percentage of students who switched from the MC to an LC answer ($\chi^2 = 4.45$; $p = 0.035$; $N = 293$; Figure 6B). There was a 23% decrease in the percentage of “D/F” students who switched from MC to LC if these students saw the histogram. In contrast, there was a 17% increase in the percent-

age of “A” students who switched from MC to LC if these students saw the histogram.

DISCUSSION

Does Clicker Participation Correlate with Grade?

In this study, we found that students earning higher grades had a higher mean participation on clicker questions (Figure 4) and less variance in participation. Students who received an “A” for the course averaged ~80% participation on clicker questions, “B” students only ~5% less, and “C” students answered 10% less frequently than “A” students. In contrast, “D/F” students averaged ~42% participation. The reasons for this drastic reduction in participation could potentially be due to several factors such as “D/F” students being absent, not participating out of fear of answering incorrectly, or a general lack of engagement.

Student participation using clickers was a significant predictor of grade. This result supports the conclusions of Jensen and Moore (2008) on the influence of student motivation on class performance. Jensen and Moore (2008) found

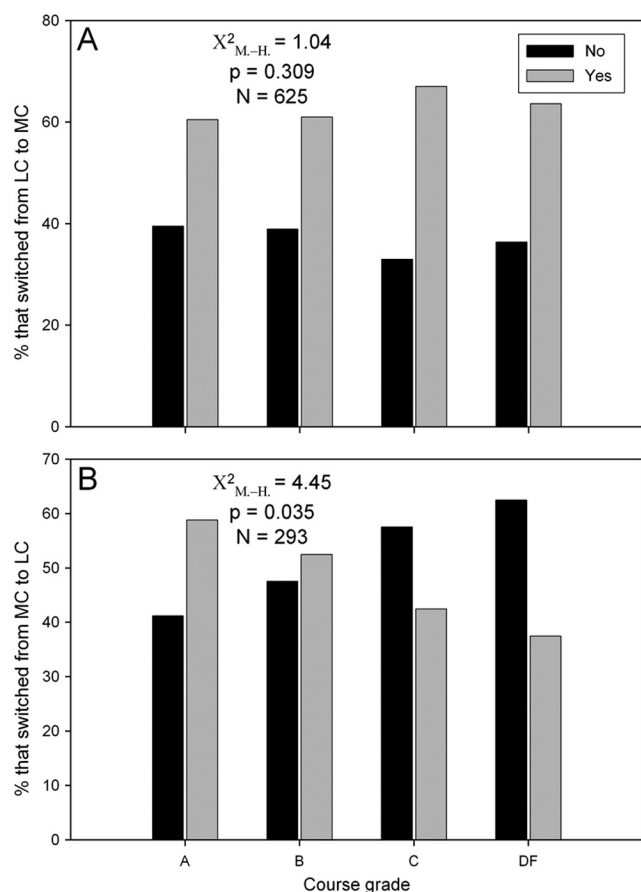


Figure 6. Percentage of students who switched their responses from LC to MC were grouped by final course grade and the percent normalized to 100% to allow for chi-square analysis. [Data for students who did not switch are not included in this analysis.] Two- and four-choice responses are combined. “No” indicates students did not see the bar graph before discussion and revoting, and “Yes” indicates students did see the bar graph. Panel A. Percentage of students who moved from LC to MC answers (No: $N_{No} = 231, N_A = 32, N_B = 113, N_C = 62, N_D = 24$; Yes: $N_{Yes} = 394, N_A = 49, N_B = 177, N_C = 126, N_D = 42$). Panel B. Percentage of students who moved from MC to LC answers (No: $N_{No} = 148, N_A = 14, N_B = 77, N_C = 42, N_D = 15$; Yes: $N_{Yes} = 145, N_A = 20, N_B = 85, N_C = 31, N_D = 9$).

that “A” students exhibit a suite of positive academic behaviors at a higher frequency than lower-scoring students, including class attendance, participation, attendance at help sessions, and completion of extra credit assignments. These results also support the idea that the lower level of clicker participation observed in “C” and “D/F” students could reflect a lower level of overall course engagement (Jensen and Moore, 2008) or academic motivation (Moore, 2007).

Using clicker participation on the 15 study questions as a way of measuring student engagement could potentially underestimate overall student engagement, because clicker participation was not a formal part of the course grade. In this class we typically tell students that consistent clicker participation could influence us to “bump up” grades that are just under a letter grade division, but correct clicks are not considered differently from incorrect clicks. Therefore,

students do not have a grade-related incentive to get the correct responses and some students may click randomly for each question without attempting to come up with a correct response. We had no way of measuring this behavior, so could not include it in the statistical analysis. However, because we anticipate that the effect of this behavior should be the same across both pre- and post-TPS votes and treatments, we suspect that it had a negligible impact on the study. In addition, the percentage of students participating in clicker questions was not significantly different across treatments and should not bias our conclusions.

Is There a Difference in the Quality of Peer Discussion If Students Are Shown the Class Bar Graph?

We found that 30% more students moved to the most common answer when they saw the graph of class responses than those in the discussion-only control group (Figure 5B). This is consistent with more of these students switching to the visibly most common answer rather than using peer discussion to move to the correct answer. If the students were changing their answers based only on peer discussion, then we would expect to see the same rate of change in those who saw the bar graph and those who did not. This may indicate that if the instructional goal of TPS is peer discussion based on the biological content of the question, the graph of student responses should not be shown before initiating student discussion because the inclination of students is to simply move to the most common answer, thus diminishing the value of the peer discussion.

In addition to the explanation that some students were biased by class responses, there are at least two other interpretations of these results. One interpretation is that seeing the most common answer provides a talking point or stimulus for more focused student discussion on why that answer was so common. Trying to identify why most of the class picked one answer could then prompt more students to switch to the correct answer. A second interpretation is that on seeing the most common, and often correct, answer, students reevaluate their own initial incorrect answer; find the flaw in their reasoning; and then switch to the correct answer on their own. Although this is a likely alternative to students merely moving to the most common answer, our data do not allow us to distinguish between these two alternatives because in most cases the correct answer was also the most common answer. Because this study was done at the end of the course, students may have also observed that the most common answer is often the correct answer, which could add some bias to their switching to the most common answer.

A second explanation of our results is that students moved to the most common answer because more of their peers had initially selected this answer and students simply changed answers based on the consensus of nearby students, but not by learning through peer instruction. Our data show one instance where this may be the case. In question 1.5 where most common was not the correct answer, students who did not see the bar graph of responses moved to the most common (but incorrect) answer. Those students did not see the graph, but they did hear the majority of their neighbors offering their incorrect answers.

Smith *et al.* (2009) tested whether learning gains are attributable to the process of peer discussion (constructivist view) or students simply being influenced by peers who know the correct answer (transmissionist view). They concluded that learning gains were due to gains in conceptual knowledge due to the process of peer discussion, regardless of whether someone in the peer discussion group knew the correct answer. Therefore, in our study, we would expect to find similar learning gains to Smith *et al.* (2009) due to peer discussion; and in our control treatment where we should only see the effect of peer discussion on movement to the correct answer, we did see a 14.5% gain (incorrect to correct answer). We attributed this to be the learning gain due to peer discussion and personal reflection. However, when students saw the bar graph this value increased to 19.3%. We are attributing this 5% difference to bias after viewing the bar graph.

Distinguishing these alternative hypotheses will require further study. One obvious follow-up to this study would include using a subset of questions of a much higher level of difficulty. This would increase the proportion of questions for which the most common answer was not the correct one. In a post hoc analysis of our data, we compared the four easiest questions with the four hardest questions and found a stronger bias on harder questions ($\chi^2 = 35.45$) than easy questions ($\chi^2 = 27.46$). Although not conclusive, this result supports our conclusions and indicates an experiment using harder questions would be a fruitful next study. In addition, lesson studies (Cerbin, 2009), recorded student discussions, student surveys, or focus groups could be used to gather data on student reflection about their own learning. In particular, we would want to ask what aspect of the process led to their changing their minds on an answer: reflection, bar graph of responses, or peer discussion. Use of isomorphic follow-up questions could help tease apart answers selected due to nearby peers conveying their answer, and whole-course peer influence. One intriguing method suggested by Lasry (2007) is to use clicker software to match students who answered questions differently for discussion. The students could then pair up or work in groups with a known set of initial responses (Lasry, 2007). This could allow explicit testing of various potential influences of peer knowledge on peer discussion.

Bearing these limitations in mind, we do feel that it is reasonable to conclude from our study that the common technique of showing the bar graph of student responses during active-learning exercises with clickers should be done with care.

Is the Effect of Seeing the Bar Graph on the Quality of Peer Discussion More Pronounced with Two-Choice or Four-Choice Questions?

Although students were more likely to switch to the most common answer regardless of the number of choices on the question, the effect was more pronounced for true/false questions with just two possible answers. A simple explanation for this observation is that students know that if one answer is not correct, then the other must be. However, for multiple-choice answers, the responses will be more distributed, and it will not be as easy to use the process of elimination to pick an answer. This observation is consistent with

research showing that multiple-choice questions have higher reliability in measuring student learning than do true/false questions (Frisbie, 1973; Ebel and Frisbie, 1991; Frisbie and Becker, 1991; Hancock *et al.*, 1993). Although true/false questions may not be inappropriate as clicker questions, the bar graph of student responses probably should not be shown between votes.

Does Student Response to Seeing the Bar Graph Correlate with Final Course Grade?

We found no correlation between course grade and the percentage of students switching from a less common to a most common answer. It seems that "A," "B," "C," and "D/F" students were influenced by whole-class peer influence at approximately equal frequencies, with approximately 30% being more likely to switch to the most common answer if they saw the graph displayed (Figure 6A). We found an interesting result when examining the percentage of students moving from a most common to a less common answer. Not surprisingly, students with higher grades tended to switch to a less common (and incorrect) answer at the lowest frequency. However, we also found that "D/F" students who did not see the display switched from a most common (and correct) answer to a less common (and incorrect) answer 62.5% of the time. Students who saw their selection displayed as the most common answer still switched to a less common answer 37.5% of the time (Figure 6B). It is possible the increased likelihood of "C" and "D/F" students' switching their answers to a less common answer reflects less confidence in their abilities and increased second guessing, or that they are sitting with peers who are also "C/D/F" students who convinced them to pick the wrong answer after peer discussion.

Our study shows that in practice it is possible to bias the quality of peer discussion by allowing the students the opportunity to see the graph of student responses. If there is a clear favorite, our study suggests that students may be biased simply by seeing the most common answer. However, when the results of an initial vote are evenly split between two or more answers, then displaying the student responses may be a valuable conversation-starter. Given these results, showing histograms during a TPS should be used judiciously.

ACKNOWLEDGMENTS

Students in the fall 2008 section of General Biology-BIO105 generously agreed to allow us to use their data by filling out an Institutional Review Board form. University of Wisconsin at La Crosse (UWL) student Kirk Gallant deciphered their signatures. We thank Mark Sandheinrich and Renee Redman for helping gather study data in their course sections. We thank Matt Evans for planting the seed of the idea that led to this study. This research was initiated as a University of Wisconsin System Faculty Scholar project by R.L.J. Bill Cerbin and Betsy Morgan provided valuable insight. K.E.P. was supported by a University of Wisconsin System Institute for Race and Ethnicity grant and the UWL College of Science and Health.

REFERENCES

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

- Armbruster, P., Patel, M., Johnson, E., and Weiss, M. (2009). Active learning and student-centered pedagogy improve student attitudes and performance in introductory biology. *CBE Life Sci. Educ.* 8, 203–213.
- Armstrong, N., Chang, S.-M., and Brickman, M. (2007). Cooperative learning in industrial-sized biology classes. *CBE Life Sci. Educ.* 6, 163–171.
- Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgment. In: *Groups, Leadership and Men*, ed. H. Guetzkow, Pittsburgh, PA: Carnegie Press, 295–303.
- Baron, R. S., Vandello, J. A., and Brunsman, B. (1996). The forgotten variable in conformity research: impact of task importance on social influence. *J. Person. Soc. Psychol.* 71, 915–927.
- Beatty, I. (2004). Transforming student learning with classroom communication systems. EDUCAUSE Center for Applied Research, 1–13.
- Bond, R., and Smith, P. B. (1996). Culture and conformity: a meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychol. Bull.* 119, 111–137.
- Bransford, J. D., Brown, A. L., and Cocking, R. R. (1999). *How People Learn: Brain, Mind, Experience and Schooling*, Washington, DC: National Academies Press.
- Caldwell, J. E. (2007). Clickers in the large classroom: current research and best-practice tips. *CBE Life Sci. Educ.* 6, 9–20.
- Cerbin, B. (2009). Assessing how students learn. Carnegie perspectives: a different way to think about teaching and learning: The Carnegie Foundation for the Advancement of Teaching. www.carnegiefoundation.org/perspectives/assessing-how-students-learn (accessed 6 November 2009).
- Chi, M.T.H., De Leeuw, N., Chiu, M.-H., and Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cogn. Sci.* 18, 439–477.
- Coleman, E. B. (1998). Using explanatory knowledge during collaborative problem solving in science. *J. Learn. Sci.* 7, 387–427.
- Coleman, E. B., Brown, A. L., and Rivkin, I. D. (1997). The effect of instructional explanations on learning from scientific texts. *J. Learn. Sci.* 6, 347–365.
- Crossgrove, K., and Curran, K. L. (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, C. H., and Mazur, E. (2001). Peer instruction: ten years of experience and results. *Am. J. Phys.* 69, 970–977.
- Crouch, C. H., Watkins, J., Fagen, A. P., and Mazur, E. (2007). Peer instruction: engaging students one-on-one, all at once. In: *Reviews in Physics Education Research*, ed. E. F. Redish, and P. Cooney, American Association of Physics Teachers, College Park, MD, 1–55.
- Ebel, R. L., and Frisbie, D. A. (1991). *Essentials of educational measurement*, Englewood Cliffs, NJ: Prentice Hall.
- 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.
- Frisbie, D. A. (1973). Multiple choice versus true-false: a comparison of reliabilities and concurrent validities. *J. Educ. Meas.* 10, 297–304.
- Frisbie, D. A., and Becker, D. F. (1991). An analysis of textbook advice about true-false tests. *Appl. Meas. Educ.* 4, 67–83.
- 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.
- Hancock, G. R., Thiede, K. W., Sax, G., and Michael, W. B. (1993). Reliability of comparably written two-option multiple-choice and true-false test items. *Educ. Psychol. Meas.* 53, 651–660.
- James, M. C. (2006). The effect of grading incentive on student discourse in Peer Instruction. *Am. J. Phys.* 74, 689–691.
- Jensen, P. A., and Moore, R. (2008). Students' behaviors, grades & perceptions in an introductory biology course. *Am. Biol. Teach.* 70, 483–487.
- Knight, J. K., and Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biol. Educ.* 4, 298–310.
- Lasry, N. (2007). Peer instruction: comparing clickers to flashcards. *Phys. Teach.* 1–4.
- Lyman, F. (1981). *The Responsive Classroom Discussion: the Inclusion of All Students*, College Park, MD: University of Maryland.
- MacManaway, L. A. (1970). Teaching methods in higher education-innovation and research. *Univ. Q.* 24, 321–329.
- Mazur, E. (1997). *Peer Instruction: A User's Manual*, Upper Saddle River, NJ: Prentice Hall.
- Moore, R. (2007). Diverse behaviors, diverse results: a motivation-based model for students' academic outcomes. In: *Diversity and the Postsecondary Experience*, ed. J. L. Higbee, D. B. Lundell, and I. M. Duranczyk, Center for Research on Developmental Education and Urban Literacy (CRDEUL), University of Minnesota: Minneapolis, MN, 129–143.
- 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. High. Educ.* 28, 457–473.
- Pollock, S. J. (2006). Transferring transformations: learning gains, student attitudes, and the impacts of multiple instructors in large lecture courses. *AIP Conf. Proc.* 818, 141–144.
- Preszler, R. W. (2009). Replacing lecture with peer-led workshops improves student learning. *CBE Life Sci. Educ.* 8, 182–192.
- Preszler, R. W., 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.
- SAS Institute (2004). *SAS/STAT 9.1 user's guide*, Cary, NC: SAS Institute.
- Slater, T. F., Prather, E. E., and Zeilik, M. (2006). Strategies for interactive engagement in large lecture science survey classes. In: *Handbook of College Science Teaching*, ed. J. J. Mintzes and W. H. Leonard, Arlington, VA: NSTA Press, 45–54.
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., and Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science* 323, 122–124.
- Walker, J. D., Cotner, S. H., Baepler, P. M., and Decker, M. D. (2008). A delicate balance: integrating active learning into a large lecture course. *CBE Life Sci. Educ.* 7, 361–367.
- Webb, N. M. (1989). Peer interaction and learning in small groups. *Int. J. Educ. Res.* 13, 21–39.
- Wood, W. B. (2004). Clickers: a teaching gimmick that works. *Dev. Cell* 7, 796–798.