

Article

Combining Peer Discussion with Instructor Explanation Increases Student Learning from In-Class Concept Questions

M.K. Smith,^{*†‡} W.B. Wood,^{*†} K. Krauter,^{*} and J.K. Knight^{*†}

^{*}Department of Molecular, Cellular, and Developmental Biology and [†]Science Education Initiative, University of Colorado at Boulder, Boulder, CO 80309

Submitted August 18, 2010; Revised December 1, 2010; Accepted December 9, 2010
Monitoring Editor: Diane Ebert-May

Use of in-class concept questions with clickers can transform an instructor-centered “transmissionist” environment to a more learner-centered constructivist classroom. To compare the effectiveness of three different approaches using clickers, pairs of similar questions were used to monitor student understanding in majors’ and nonmajors’ genetics courses. After answering the first question individually, students participated in peer discussion only, listened to an instructor explanation only, or engaged in peer discussion followed by instructor explanation, before answering a second question individually. Our results show that the combination of peer discussion followed by instructor explanation improved average student performance substantially when compared with either alone. When gains in learning were analyzed for three ability groups of students (weak, medium, and strong, based on overall clicker performance), all groups benefited most from the combination approach, suggesting that peer discussion and instructor explanation are synergistic in helping students. However, this analysis also revealed that, for the nonmajors, the gains of weak performers using the combination approach were only slightly better than their gains using instructor explanation alone. In contrast, the strong performers in both courses were not helped by the instructor-only approach, emphasizing the importance of peer discussion, even among top-performing students.

INTRODUCTION

Active-learning activities can significantly increase student learning in biology courses (Udovic *et al.*, 2002; Kitchen *et al.*, 2003; Knight and Wood, 2005; Freeman *et al.*, 2007; Walker *et al.*, 2008). Among the many kinds of such activities that are practical in large lecture classrooms, in-class concept ques-

tions using personal response systems or “clickers” have received recent attention (e.g. Wood, 2004; Caldwell, 2007). Typically, instructors pose multiple-choice questions requiring application of a recently presented concept at several time points during a class, and students record their answers using clickers. In addition to breaking up lectures into smaller chunks, concept questions provide students with opportunities to practice solving problems and monitor their understanding during class. Recent work in cognitive psychology has shown that frequent assessment of students in this manner has a powerful impact on both learning and retention (reviewed in Roediger *et al.*, 2010).

In connection with in-class concept questions, instructors often use an approach called peer instruction, which encourages students to verbalize their thinking and interact with their peers to arrive at an answer (Mazur, 1997). In one commonly used mode, students first answer a concept question individually, discuss the question with their peers, and then revote before the answer to the question is revealed. The instructor then explains the question and often shows a

DOI: 10.1187/cbe.10-08-0101

Address correspondence to: Michelle K. Smith (micsmith@u.washington.edu).

[‡]Present address: Department of Genome Sciences, University of Washington, Seattle, WA 98195.

© 2011 M. K. Smith *et al.* CBE—Life Sciences Education © 2011 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution-Noncommercial-Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB[®]” and “The American Society for Cell Biology[®]” are registered trademarks of The American Society for Cell Biology.

Table 1. Demographic information on students who participated in this study

| Category | Majors' course | Nonmajors' course |
|---------------------------------------|---|---|
| Gender | 41% female, 59% male | 66% female, 34% male |
| Year in college | 7% freshman, 31% sophomore, 29% junior, 26% senior, 7% other | 36% freshman, 37% sophomore, 11% junior, 13% senior, 3% other |
| Major | 55% biology; 98% indicated they hoped to pursue a career related to science | 10% biology ^a |
| Grade distribution in genetics course | 26% A, 39% B, 22% C 9% D, 4% F | 37% A, 36% B, 23% C, 3% D, 0% F |

^aThere are three biology majors at CU-Boulder: Molecular, Cellular, and Developmental Biology (MCDB); Ecology and Evolutionary Biology (EBIO); and Integrative Physiology (IPHY). These students are EBIO and IPHY majors.

histogram of the student responses, which gives both instructors and students immediate feedback on how well a concept is understood.

Many instructors report that the frequency of correct answers increases after peer discussion (Mazur, 1997; Crouch and Mazur, 2001; Knight and Wood, 2005; Smith *et al.*, 2009). Two alternative hypotheses could explain this observation: 1) active engagement of students during discussion with peers leads to increased conceptual understanding, resulting in improved performance on the revote, or 2) students do not necessarily learn from the discussion, but simply choose the answer most strongly advocated by neighbors they perceive to be knowledgeable.

In a previous study (Smith *et al.*, 2009), strong support was obtained for the first hypothesis using matched pairs of in-class concept questions that addressed the same concept and required similar reasoning but had a different story line. Students answered the first question of a pair (Q1) individually. After a few minutes spent discussing their responses in small groups, they revoted on Q1. Students then answered a second question (Q2) individually, and only then were the answers and the histograms for both questions revealed and discussed. Subsequent tracking of student responses using the clicker software showed that students who changed their answers to Q1 from incorrect to correct after discussion performed on average much better on Q2 than students who did not change their answers. Moreover, on the more difficult questions, performance on both Q1 after discussion and Q2 increased markedly, even for groups in which no student initially answered Q1 correctly, indicating that the process of discussion itself rather than the influence of knowledgeable peers could lead students to increased understanding.

Although this study demonstrated that peer discussion had a positive effect on student learning, it did not attempt to compare peer discussion with explanation from the instructor as an alternative activity between Q1 and Q2. In informal discussions with instructors, we learned that some instructors skip peer discussion, believing that their explanation of a clicker question answer will be clearer, more efficient, and more informative than what students are likely to hear in conversations with each other. However, the constructivist viewpoint supported by the above study of Smith *et al.* (2009) predicts that the process of verbalization and discussion could promote understanding more effectively than even a clear instructor explanation. In addition, grappling with a question in discussions with peers could enhance the learning value of a subsequent explanation by the instructor (Schwartz and Bransford, 1998).

To explore the merits of these alternative views, we applied a modification of our earlier protocol using matched pairs of questions (referred to in that study as isomorphic questions; Smith *et al.*, 2009) to ask which of the following three presentation modes leads to the greatest improvement in student performance: having a peer discussion, listening to an instructor explanation, or engaging in peer discussion followed by an instructor explanation (the combination approach). We evaluated the effects on student learning gains in two classes, genetics for majors and nonmajors, as well as for three different ability groups of students classified as strong, medium, and weak clicker performers.

METHODS

Student Demographics

This study was conducted in an undergraduate introductory genetics course required for majors (Fall semester of 2008) and a genetics course for nonmajors (Fall semester of 2009) (student demographics shown in Table 1). These courses were taught in the Department of Molecular, Cellular, and Developmental Biology (MCDB) at the University of Colorado, Boulder, by two of the authors: K.K. (majors) and J.K.K. (nonmajors). Both courses met for three 50-min sessions per week, and student grades in both courses were based on a similar distribution of points (Table 2).

Instructional Modes and Experimental Protocols

In both courses, an average of four in-class concept questions were asked per class, and approximately half the class periods included matched-pair questions that were used in this study. Even though all the in-class concept questions were awarded only participation points, students had an incentive to do

Table 2. Distribution of grading points in the majors' and nonmajors' genetics courses

| | Majors (%) | Nonmajors (%) |
|--|------------|---------------|
| Exams | 71 | 61 |
| Homework | 17 | 11 |
| Clicker participation | 10 | 12 |
| Additional participation (surveys, reflections) | 2 | 5 |
| Group project | N/A | 11 |

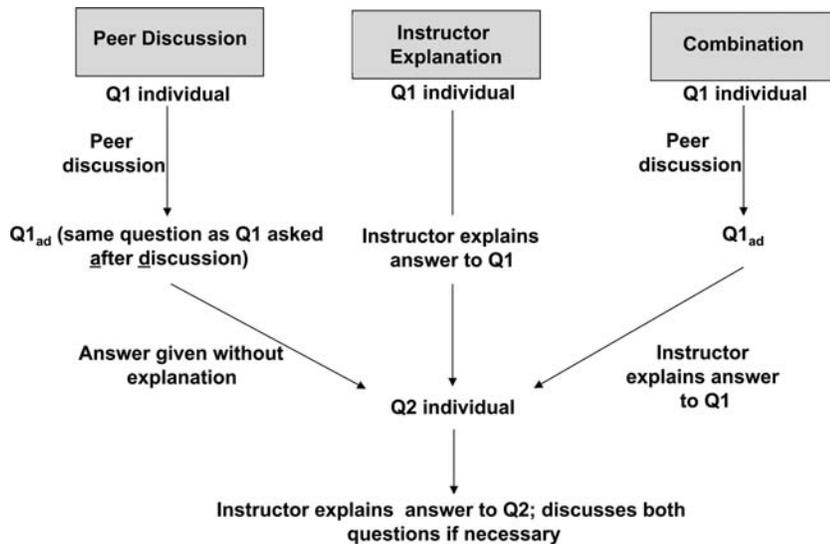


Figure 1. Flowchart of the experimental design for investigating whether students learn more from in-class concept questions by peer discussion, listening to an instructor explanation, or participating in a combination of peer discussion followed by instructor explanation. These variations of the protocol are referred to as modes in the text. Q1: one question of a matched pair was voted on individually; Q1_{ad}: the same question was voted on again after peer discussion; Q2: the second very similar question was voted on individually. See *Methods* for further detail.

well on these questions because they were told that clicker questions gave them practice for the exams.

To determine whether engaging in peer discussion, listening to an instructor explanation, or participating in a combination of peer discussion followed by instructor explanation was more effective for student learning, we followed each of three different modes of the experimental protocol outlined in Figure 1, using matched pairs of questions that test genetics concepts as described in the *Introduction*. Both courses have similar learning objectives and questions on similar topics, but the question pairs used were different because of the higher level of detail appropriate for the majors' course (examples shown in Figure 2).

Our protocol included three types of questions: Q1, Q1_{ad}, and Q2 (Figure 1). In all three modes, students answered Q1 individually to provide a measure of student understanding after listening to a lecture on the topic. The three modes were as follows:

In the peer-discussion mode, students revoted on Q1 after discussion (Q1_{ad}). After recording their vote, they were told the correct answer to the question, but no additional explanation was given.

In the instructor mode, after students had answered Q1 individually, the instructor asked the students to volunteer their reasons for selecting specific answers, explained the solution to Q1, and answered any student questions.

For the combination mode, after students answered Q1 individually, they discussed the question with their neighbors and then voted on the same question again (Q1_{ad}), just as in the peer-discussion mode. The instructor then asked the students to volunteer their reasons for selecting specific answers, explained the solution to Q1, and answered any student questions, as in the instructor-explanation mode.

In all three modes, students then individually voted on Q2.

After all of the Q2 votes were recorded, the instructor explained the solutions to Q1 (for the peer-discussion mode) and Q2. Histograms of student responses to Q1 and Q2 were shown only after the Q2 vote, because showing histogram results can bias student responses and influence the student discussion (Perez *et al.*, 2010).

To compare these instructional modes in a normal classroom setting, there were no time limits placed on the instructor explanations or student voting. Both instructors generally let student voting continue until 75–80% of the students had recorded their vote, encouraged the remaining students to vote, and then stopped the voting 10–20 s later. Consequently, mean amounts of time devoted to consideration of Q1 for the different modes varied (Table 3). For both the majors' and the nonmajors' courses, more time on average was spent considering answers to Q1 in the combination mode than in the peer-discussion or instructor-explanation modes, as might be expected (implications of these differences are explored in the *Discussion*).

All three modes of the protocol were used in the majors course, but only the instructor-explanation and combination modes were used in the nonmajors' course.

Description of Matched Pair Questions

To minimize any bias toward writing an easier Q2 question, the questions in each pair were randomly assigned to be Q1/Q1_{ad}, or Q2 after they were written (Smith *et al.*, 2009). Which of the three modes of presentation to use was also randomly determined for each question pair. Q1/Q1_{ad} and Q2 were assigned to a mode of presentation and inserted into the slide presentations shortly before class to minimize the possibility of altering the lecture to favor one mode of presentation over another.

Both instructors agreed that questions where the individual Q1 vote was greater than 80% correct were insufficiently challenging and left little opportunity for gains in learning; these questions were not included in this study. Although we intended all questions to be challenging, one peer-discussion and two combination questions had Q1 scores of >80% correct in the majors' course. In the nonmajors' course, three instructor-explanation questions had Q1 scores of >80% correct.

After the course was completed, all the question pairs used were judged for similarity by two independent reviewers, who did not have access to the student performance results.

Majors Q1/Q1_{ad}: UNC has hired you to generate gold bear mascots. You mutate a bunch of bears and get 4 that are gold because of recessive mutations. You carry out crosses between all the bears (all bears can mate with each other, don't worry about mascot sex) and produce the following results where a "-" means mutant gold bears were produced and a "+" means wild-type brown bears were produced.

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | | - | + | ? |
| 2 | | | + | - |
| 3 | | | | + |
| 4 | | | | |

- Which statement is correct?
- A. 1 crossed with 4 will produce a gold bear
 - B. 1 crossed with 4 will produce a brown bear
 - C. You need more information to answer this question

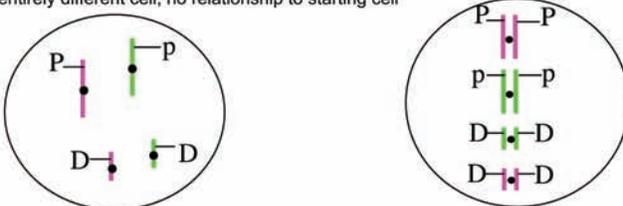
Majors Q2: CSU has hired you to generate green ram mascots. You mutate a bunch of rams and get 4 that are green because of recessive mutations. You carry out crosses between all the rams (all rams can mate with each other, don't worry about mascot sex) and produce the following results where a "-" means mutant green rams were produced and a "+" means wild-type white rams were produced.

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | | + | ? | - |
| 2 | | | - | + |
| 3 | | | | + |
| 4 | | | | |

- Which statement is correct?
- A. 1 crossed with 3 will produce a green ram
 - B. 1 crossed with 3 will produce a white ram
 - C. You need more information to answer this question

Non-majors Q1/Q1_{ad}: The cell on the left is a normal cell in the human body, with 4 of 46 chromosomes shown. The cell to the right is an example of:

- a. This cell in Interphase
- b. This cell in Metaphase of mitosis
- c. This cell in Metaphase of meiosis 1
- d. A daughter of this cell in Metaphase of meiosis II
- e. An entirely different cell; no relationship to starting cell



Non-majors Q2: The cell on the left is a normal cell in the human body, with 4 of 46 chromosomes shown. The cell to the right is an example of:

- a. This cell in Interphase
- b. This cell in Metaphase of mitosis
- c. This cell in Metaphase of meiosis 1
- d. A daughter of this cell in Metaphase of meiosis II
- e. An entirely different cell; no relationship to starting cell

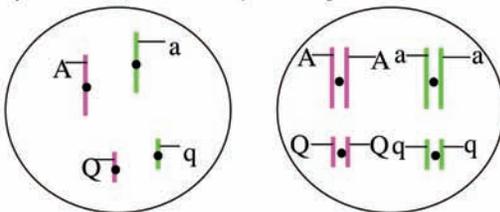


Figure 2. Examples of the Q1/Q1_{ad} and Q2 question pairs used in this study. The correct answers are underlined.

The reviewers were familiar with the content of the genetics courses and had participated in an earlier study that used matched pairs of questions to measure the benefits of peer discussion (Smith *et al.*, 2009). These reviewers were asked to judge whether they thought the question pairs in this study

were testing the same concept. Data from five question pairs, three from the majors' course and two from the nonmajors' course, were removed from the data set, because two independent reviewers judged them as not testing identical concepts. Individual responses were also removed from the data

Table 3. Average elapsed times between the end of individual Q1 and the start of individual Q2 for the three modes of discussing Q1

| Mode of discussing Q1 | Majors ^a | Nonmajors ^a |
|------------------------|----------------------|------------------------|
| Peer discussion | 2 min 54 s (17 s) | N/A ^b |
| Instructor explanation | 1 min 54 s (25 s) | 3 min 19 s (45 s) |
| Combination | 4 min 42 s (53 s) | 5 min 5 s (60 s) |

^aThe SEM is in parentheses.

^bThe peer discussion alone mode was not used in the nonmajors' course (see text).

set if a student did not answer all questions in a question set (e.g., answered Q1 and Q1_{ad} but not Q2).

The remaining 32 question pairs were also rated for cognitive level according to Bloom's taxonomy (Bloom and Krathwohl, 1956) by two independent reviewers not associated with the course who are experts at these rankings (Crowe *et al.*, 2008). The raters were given all 64 questions in a random order and were not told which questions were matched-pair sets. Both raters independently determined that the average level of the questions was 3 (application level). The two raters concluded that 87% and 81%, respectively, of the Q1–Q2 pairs were at the same Bloom's level. For the question pairs that did not match, 60% of the time the raters concluded that Q2 was at a higher level than Q1.

The data set included responses from 150 students in the majors class and 62 students in the nonmajors class. These students answered at least one complete set of questions in each of the different modes. Table 4 shows the mean number of questions answered for each protocol mode.

Data Analysis

The change in learning between question pairs was computed for each individual student using a modified version of the Hake normalized gain formula (Hake, 1998) known as normalized change <c> (Marx and Cummings, 2007). Normalized change values provide a measure of how much a student's performance increases compared with that individual's maximum possible increase. When calculating the mean normalized change between Q1 and Q2 over all question pairs for a given student, the following formula was used when an individual's mean Q2 score was higher than the mean Q1 score (most cases): $\langle c \rangle = 100(\text{mean Q2} - \text{mean Q1}) / (100 - \text{mean Q1})$. Alternatively, if an individual's mean Q1 score was

Table 4. Mean number of questions answered for each protocol mode

| Mode of discussing Q1 | Majors | Nonmajors |
|------------------------|--------------|------------------|
| Peer discussion | 5.8 out of 7 | N/A ^a |
| Instructor explanation | 5.3 out of 6 | 4.0 out of 5 |
| Combination | 4.0 out of 5 | 7.1 out of 9 |

^aThe peer discussion alone mode was not used in the nonmajors' course (see text).

higher than the mean Q2 score, $\langle c \rangle = 100((\text{Q2} - \text{Q1}) / \text{Q1})$, was used. In cases where an individual's mean Q1 score and the mean Q2 score equaled either 100 or 0, the response for that student was removed from the data set, because otherwise <c> would be recorded as 0. Significant differences between mean <c> values between two populations cannot be determined because they are nonlinear computed quantities that are not normally distributed. Instead, the standard error measurements on reported <c> values are used to provide a coarse depiction of the spread of values (Marx and Cummings, 2007).

All statistical analyses were performed with SPSS (SPSS, Chicago, IL) or Excel (Microsoft, Redmond, WA). Item discrimination values (D) were calculated by rank, ordering students by their overall Q1 percent correct score. The top 27% and the bottom 27% of students in the majors' and nonmajors' courses were compared for this analysis. For each Q1 question, the following formula was used: $D = (R_U - R_L) / (1/2T)$. R_U is the number of students in the upper group who answered correctly, R_L is the number of students in the lower group who answered correctly, and T is the total number of students included in the item analysis (Gronlund, 1976). The average item discrimination values for Q1 questions were then calculated for each protocol mode in the majors' and nonmajors' courses.

Institutional Review Board Protocols

Approval to evaluate student clicker responses (exempt status, Protocol No. 0108.9) and end-of-year survey responses (expedited status, Protocol No. 0603.08) was granted by the Institutional Review Board, University of Colorado, Boulder.

RESULTS

Q1 Questions Have Equivalent Difficulty and Adequate Item Discrimination for Question Pairs Administered in Each of the Three Modes

The mean percentages of correct individual Q1 answers were not statistically different between the three different protocol modes for the majors (Figure 3A, repeated measures analysis of variance, $p > 0.05$). Similarly, for the nonmajors, the percentages of correct individual Q1 answers were not statistically different between the instructor-explanation and combination modes (Figure 3B, paired *t*-test, $p > 0.05$). Also, the average item discrimination values (D) for the Q1 questions were greater than 0.3 for all protocol modes in both the majors' and nonmajors' courses (Table 5). Questions with D values above 0.3 are generally considered good discriminators of the top and bottom students (Ebel, 1965).

The Learning Gains between Q1 and Q1_{ad} Were Similar in Both Modes That Involved Peer Discussion

An initial measure of learning through peer discussion was calculated by recording mean student performance on individual Q1s and the same questions after discussion (Q1_{ad}) (Figure 3). In all cases, students' mean performance on Q1_{ad} was significantly higher than on Q1 (dependent *t*-test, $p < 0.05$). In addition, we calculated the mean normalized change (<c>) between Q1 and Q1_{ad} for each individual student. In

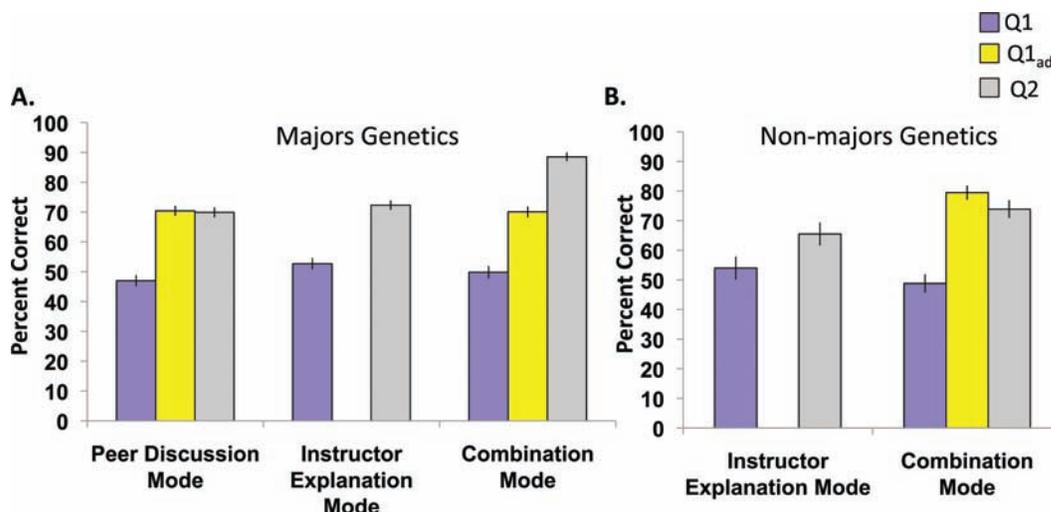


Figure 3. Effects of three different modes of discussing Q1 on percentage of Q1_{ad} and Q2 correct answers in the majors' (A) and nonmajors' (B) genetics courses. Performance results were averaged for each individual before computing the means shown. The SEMs are shown with error bars.

the majors' course, this value was 41.5% (± 3.4) for the peer-discussion mode and 37.2% (± 3.4) for the combination mode. The similarity of these values suggests that peer discussion resulted in similar performance improvement for both these modes. For the nonmajors, the mean $\langle c \rangle$ between Q1 and Q1_{ad} for the combination mode was somewhat higher at 56.9% ($\pm 5.5\%$).

The Combination Mode Led to Larger Learning Gains between Q1 and Q2 Than Either Peer Discussion or Instructor Explanation Alone

In the majors' course, when all three modes of the protocol were compared for student performance on Q1 and Q2, the mean percentage of correct answers was higher for Q2, indicating that performance improved in all three modes (Figure 3, dependent t -test, $p < 0.05$ in all cases). Similarly in the nonmajors' course, the mean percentage of correct answers was significantly higher for Q2 than for Q1, indicating that performance improved in both the instructor-explanation and combination modes (Figure 3, dependent t -test, $p < 0.05$ in all cases).

Two principal findings from comparisons of learning gains are presented in Figure 4, which shows the Q1-to-Q2 mean $\langle c \rangle$ values for each mode in both the majors' and nonmajors'

genetics courses. First, in the majors' course, the peer-discussion and instructor-explanation modes resulted in similar mean $\langle c \rangle$ values, suggesting that each of these modes alone is equally effective. Second, in both courses, the combination mode resulted in strikingly higher $\langle c \rangle$ values than either of the other modes alone.

The Combination Mode Results in the Largest Gain in Learning for All Ability Levels of Students

To determine whether a certain instructional mode is better for students who tend to do well or poorly on in-class concept questions, mean Q1 percent correct scores for all instructional modes were calculated for each student. Then the students in each course were divided into three groups based on these scores, in which weak, medium, and strong clicker performers were designated as having mean Q1 scores of $< 33.3\%$, $33.3\text{--}66.6\%$, and $> 66.6\%$, respectively. Table 6 shows the percentage of students who fell into each category. The majority of students in both courses fell into the medium clicker performer category.

Figure 5 shows the learning gains represented by average Q1-to-Q2 $\langle c \rangle$ scores for the weak, medium, and strong clicker performers for all three instructional modes in both courses. In the majors' course (Figure 5A), the combination mode was clearly most effective for all three groups of students. For the weak and medium groups, the peer-discussion and instructor-explanation modes appeared equally effective, whereas for the strong group, the instructor mode appeared least effective.

Similar trends were seen in the nonmajors' course (Figure 5B). Namely, the combination mode was most effective for all three groups of students, except for the weak performers, for whom the gains for the instructor-explanation and combination modes were similar. As was true with the majors, the instructor-explanation mode was least effective for the strong students.

Table 5. Mean item discrimination (D) values for Q1 questions in the different protocol modes

| Method of discussing Q1 | Majors | Nonmajors |
|-------------------------|--------|------------------|
| Peer discussion | 0.42 | N/A ^a |
| Instructor explanation | 0.39 | 0.39 |
| Combination | 0.37 | 0.55 |

^aThe peer discussion alone mode was not used in the nonmajors' course (see text).

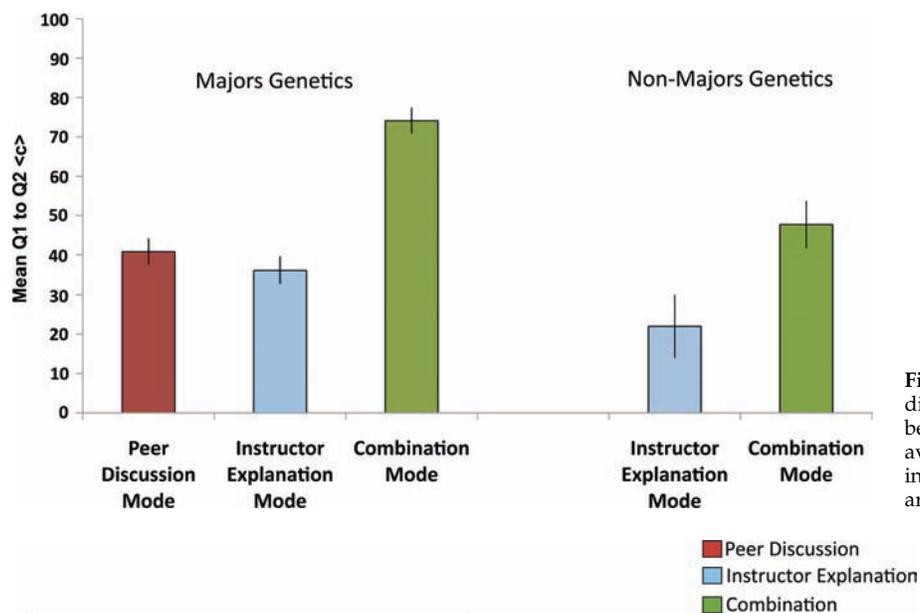


Figure 4. Effects of three different modes of discussing Q1 on the normalized change <c> between Q1 and Q2. Performance results were averaged for each individual before computing the means shown (see *Methods*). The SEMs are shown with error bars.

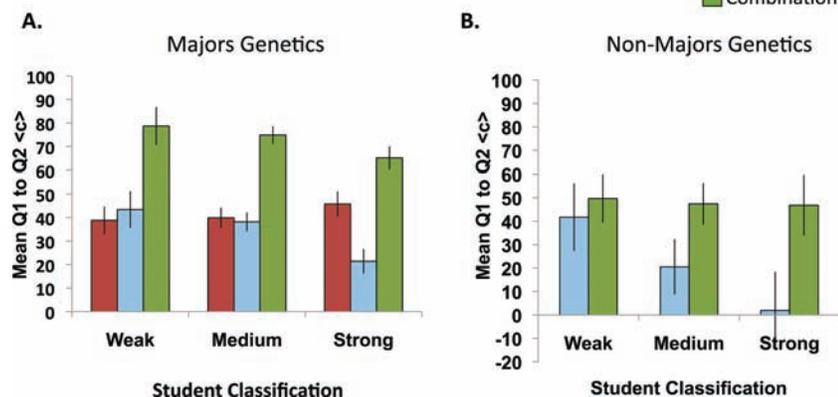


Figure 5. Effects of three different modes of discussing Q1 on mean for students in the majors' (A) and nonmajors' (B) genetics courses classified as weak, medium, or strong clicker performers based on their overall mean Q1 score (see *Results*). SEM are shown with error bars.

DISCUSSION

Summary of Results

Our results show that genetics students, both majors and nonmajors, learn from in-class concept questions whether the mode of administration comprises peer discussion alone, instructor explanation alone, or a combination mode in which peer discussion is followed by instructor explanation (Figure 3). However, the combination mode results in substantially higher learning gains compared with either the peer-discussion or instructor-explanation modes, as measured by the normalized change <c> in scores between Q1 and Q2 (Figure 4). Analysis of the results for three ability groups of

students, designated weak, medium, and strong based on mean Q1 scores, showed that the combination mode was most effective for all three groups in both the majors' and the nonmajors' courses (Figure 5).

Strikingly, the strong clicker performers in both classes showed the smallest learning gains when the instructor-explanation mode was used (Figure 5). We hypothesize that discussing questions with peers in either the peer-discussion mode or the combination mode keeps the strong clicker performers engaged with the material. Without this element in the instructor mode, the strong students may pay less attention to the subsequent question, Q2. These results are in agreement with a study that compared overall student learning gains in introductory physics courses taught using traditional lecturing or interactive engagement (Beichner and Saul, 2003). In this study, the stronger students learned more in the interactive courses, possibly because they were cementing their own understanding by helping their peers. Our results support the conclusions of these authors that interactive approaches such as peer discussion benefit the high-achieving students.

We see differences between students in the majors' and nonmajors' genetics courses with respect to the weak clicker performers. In the majors' course, the weak students show

Table 6. Distribution of students classified into the weak, moderate, and strong clicker performer groups in the majors' and nonmajors' genetics courses

| | Majors (%) | Nonmajors (%) |
|-----------------------------|------------|---------------|
| Weak clicker performers | 18 | 19 |
| Moderate clicker performers | 63 | 57 |
| Strong clicker performers | 19 | 24 |

substantially larger learning gains with the combination mode than with either of the other two modes (Figure 5A). However, for weak students in the nonmajors' course, the combination mode is only slightly more effective than the instructor-explanation mode (Figure 5B). One likely reason for this difference is that nonmajors were less inclined to regard their peers as learning resources. Several lines of support for this idea come from a previous study in which behaviors and motivation levels of nonmajor genetics students were measured (Knight and Smith, 2010). Observations of these students revealed that they were more likely than majors to ask an instructor rather than peers for help when working on group activities. Nonmajors in this study also studied outside of class significantly less than did majors, consistent with lower levels of motivation. These factors may combine to generate an environment for nonmajors in which the weaker students are less inclined to participate in peer discussion, and thus do not benefit as much as other groups.

Our data from the majors' genetics course show that the peer-discussion and instructor-explanation modes result in similar learning gains, at least for the weak and medium clicker performers (Figures 4 and 5A). Peer discussion has benefits over listening to an instructor, such as breaking up the monotony of lecture and giving students a chance to practice putting their thoughts into words (Mazur, 1997; Smith *et al.* 2009). However, in our experience, many students report that peer discussion without any instructor explanation or feedback can be frustrating.

Why Is the Combination of Peer Discussion Followed by Instructor Explanation So Effective for Student Learning?

The effectiveness of the combination mode is consistent with previous findings in cognitive psychology, showing that student engagement in a learning activity such as answering questions predisposes them to learn from a subsequent lecture (Schwartz and Bransford, 1998). During peer discussion, students engage with the material by sharing their ideas with others. In short, students are figuring out what they understand and what they have questions about. The instructor explanation immediately following peer discussion in our protocol corresponds to the subsequent lecture in the Schwartz and Bransford (1998) study. Additional studies have shown that feedback to students, which allows them to gauge their current understanding of a topic, can have a positive impact on their future performance. Feedback is especially helpful when it includes a statement of the correct answer and information on why it is correct (reviewed in Roediger *et al.*, 2010). In both the instructor-explanation and combination modes used in this study, students received extensive feedback of this nature from their instructor, as well as explanations of why other answers were incorrect.

In the combination mode, students spent on average about 2–3 min more total time engaging with Q1 than in either of the other two modes (Table 3). Time on task was not strictly controlled in our study because we wanted to compare these modes in a normal classroom setting without imposing time limits on the instructor. However, several considerations argue that time on task alone cannot account for the superior effectiveness of the combination mode. From our experience in the classroom, in agreement with published guides to best

practices with clickers (Caldwell, 2007), useful peer discussion following a question is generally limited to 2–3 min, after which most of the students turn to conversations on other matters or personal pursuits, such as email or texting. Therefore, simply adding time to peer discussion would be highly unlikely to increase the effectiveness of this mode significantly. Consistent with this view is a recent physics education study, in which students individually answered a clicker question and then for the next minute discussed the question with their peers, reflected on their answers silently, or were distracted by a cartoon (Lasry *et al.*, 2009). When the students then voted on the question again, the percent change in performance was highest when students engaged in peer discussion, suggesting that the benefit of this activity is not simply to provide additional time considering the question. Based on these arguments, the substantially increased learning gains that result from adding instructor explanation after peer discussion are highly unlikely to be attributed merely to the increased time on task.

Could a longer and more detailed instructor explanation following administration of Q1 have increased the effectiveness of this mode alone to the level of the combination mode? Although we did not do the experiment, three considerations suggest that this possibility is unlikely as well. The first is the size of the effect; the combination mode was on average approximately twice as effective in promoting learning gains as the instructor-explanation mode, in both courses. Adding 2 or 3 min of instructor explanation to an already complete explanation is unlikely to have produced such a doubling. Second, the findings of Schwartz and Bransford (1998) suggest that, after grappling with a question, students are primed to gain more from a subsequent lecture. These results suggest an apparent synergy, which we have also observed, between peer discussion and instructor explanation in the combination mode. Third, student surveys also indicate that they perceive the combination mode to be synergistic. On end-of-course surveys in both courses ($n = 122$ major respondents, $n = 45$ nonmajor respondents), students were asked to indicate agreement or disagreement with the following statement: "Having a discussion with my neighbors prepares me to listen to the instructor's explanation." Sixty-four percent of the majors and 84% of the nonmajors agreed, and when asked to explain why, several students described how peer discussion helps prepare them to learn. Two such descriptions follow:

It gets me thinking about the topic before [the instructor's] lecture, rather than just passively listening to what he has to say – I am already engaged.

Discussion helps get the ideas and thoughts flowing, which makes what [the instructor] says more concrete.

The class time required for administration and discussion of concept questions, especially using the combination mode, may seem daunting to some instructors. However, several studies have shown the value of modifying course structure so as to place more responsibility on students for learning factual material outside of class, thereby freeing class time for active-learning activities, such as clicker questions and discussion (reviewed in Wood, 2009). In the courses described here, the instructors focused primarily on conceptual understanding in class rather than transmission of detailed factual

knowledge. In addition, students practiced general skills and application of concepts through regular online homework assignments outside of class. These modifications allowed the instructors to require mastery of basic content while still leaving time for in-class active-learning activities.

CONCLUSION AND FUTURE DIRECTIONS

Our research further defines best practices for using in-class concept questions and clickers. From previous work, we know that active engagement of students during peer discussion leads to improved performance (Smith *et al.*, 2009). The results presented here show that, in two different courses, the largest gains in student performance occur when peer discussion is immediately followed by instructor explanation. This combination mode is probably so effective because it combines student engagement through peer discussion with instructor feedback. Qualitative studies on the content of student discussions during peer interaction, currently in progress, should help to better understand the benefits of this mode of clicker use.

Although our results indicate that the combination mode is better for in-class performance than the single modes tested, we still do not know which modes of instruction best promote retention of material. Following the evidence from cognitive psychology studies (Schwartz and Bransford, 1998; Roediger *et al.*, 2010), we would predict that peer discussion immediately followed by instructor explanation should enhance not only short-term learning, but also retention as well. Longitudinal studies are needed to explore this prediction.

ACKNOWLEDGMENTS

We are grateful to the Science Education Initiative at the University of Colorado, Boulder, for full support of M.K.S and partial support of J.K.K during this project. Thanks to Alison Crowe and Mary Pat Wenderoth for assessing the Bloom's levels of the question pairs. We also thank Carl Wieman and Wendy Adams for intellectual support and comments on the manuscript and Tin Tin Su for help with the experimental design.

REFERENCES

- Beichner RJ, Saul JM (2003). Introduction to the SCALE-UP (Student-Centered Activities for Large Enrollment Undergraduate Programs) Project. Proceedings of the International School of Physics "Enrico Fermi," Varenna, Italy.
- Bloom BS, Krathwohl DR (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain, New York: Longmans.
- Caldwell JE (2007). Clickers in the large classroom: current research and best-practice tips. *CBE Life Sci Educ* 6, 9–20.
- Crouch CH, Mazur E (2001). Peer instruction: ten years of experience and results. *Am J Phys* 69, 970–977.
- 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.
- Ebel RL (1965). *Measuring Educational Achievement*, Englewood Cliffs, NJ: Prentice-Hall.
- Freeman S, O'Connor E, Parks JW, Cunningham M, Hurley D, Haak D, Dirks C, Wenderoth MP (2007). Prescribed active learning increases performance in introductory biology. *CBE Life Sci Educ* 6, 132–139.
- Gronlund NE (1976). *Measurement and Evaluation in Teaching*, New York: Macmillan.
- Hake RR (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.
- Kitchen E, Bell JD, Reeve S, Sudweeks RR, Bradshaw WS (2003). Teaching cell biology in the large-enrollment classroom: methods to promote analytical thinking and assessment of their effectiveness. *Cell Biol Educ* 2, 180–194.
- Knight JK, Smith MK (2010). Different but equal? How nonmajors and majors approach and learn genetics. *CBE Life Sci Educ* 9, 33–44.
- Knight JK, Wood WB (2005). Teaching more by lecturing less. *Cell Biol Educ* 4, 298–310.
- Lasry N, Charles E, Whittaker C, Lautman M (2009). When talking is better than staying quiet. *PERC* 1179, 181–184.
- Marx J, Cummings K (2007). Normalized change. *Am J Phys* 1, 87–91.
- Mazur E (1997). *Peer Instruction: A User's Manual*, Saddle River, NJ: Prentice Hall.
- 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.
- Roediger HL, Agarwal PK, Kang SHK, Marsh EJ (2010). Benefits of testing memory: best practices and boundary conditions. In: *New Frontiers in Applied Memory*, eds. GM Davies and DB Wright, Brighton, UK: Psychology Press, 13–49.
- Schwartz DL, Bransford JD (1998). A time for telling. *Cogn Instr* 16, 475–522.
- 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.
- Udovic D, Morris D, Dickman A, Postlethwait J, Wetherwax P (2002). Workshop biology: demonstrating the effectiveness of active learning in an introductory biology course. *Bioscience* 52, 272–281.
- Walker JD, Cotner SH, Baepler PM, Decker MD (2008). A delicate balance: integrating active learning into a large lecture course. *CBE Life Sci Educ* 7, 361–367.
- Wood WB (2004). Clickers: a teaching gimmick that works. *Dev Cell* 7, 796–798.
- Wood WB (2009). Innovations in teaching undergraduate biology and why we need them. *Annu Rev Cell Dev Biol* 25, 93–112.