CBE—Life Sciences Education
Vol. 8, 203–213, Fall 2009

Article

Active Learning and Student-centered Pedagogy Improve Student Attitudes and Performance in Introductory Biology

Peter Armbruster,* Maya Patel,*† Erika Johnson,* and Martha Weiss*

*Department of Biology, Georgetown University, Washington, DC 20057; and †Department of Education, Cornell University, Ithaca, NY 14853

Submitted March 27, 2009; Accepted May 28, 2009
Monitoring Editor: Debra Tomanek

We describe the development and implementation of an instructional design that focused on bringing multiple forms of active learning and student-centered pedagogies to a one-semester, undergraduate introductory biology course for both majors and nonmajors. Our course redesign consisted of three major elements: 1) reordering the presentation of the course content in an attempt to teach specific content within the context of broad conceptual themes, 2) incorporating active and problem-based learning into every lecture, and 3) adopting strategies to create a more student-centered learning environment. Assessment of our instructional design consisted of a student survey and comparison of final exam performance across 3 years—1 year before our course redesign was implemented (2006) and during two successive years of implementation (2007 and 2008). The course restructuring led to significant improvement of self-reported student engagement and satisfaction and increased academic performance. We discuss the successes and ongoing challenges of our course restructuring and consider issues relevant to institutional change.

INTRODUCTION

The traditional lecture format of most large introductory science courses presents many challenges to both teaching and learning. Although a traditional lecture course may be effective for efficiently disseminating a large body of content to a large number of students, these one-way exchanges often promote passive and superficial learning (Bransford *et al.*, 2000) and fail to stimulate student motivation, confidence, and enthusiasm (Weimer, 2002). As a consequence, the traditional lecture model can often lead to students completing their undergraduate education without skills that are important for professional success (National Research Council [NRC], 2007; also see Wright and Boggs, 2002, p. 151). Over the past two decades, a series of influential reports and articles have called attention to the need for changes in approaches to undergraduate science education in ways that promote meaningful learning, problem solving, and critical thinking for a diversity of students (American Association for the Advancement of Science, 1989; Boyer, 1998; NRC, 1999, 2003, 2007; Handelsman *et al.*, 2004, 2007; Project Kaleidoscope, 2006). This need is particularly acute at the introductory level, where a major "leak in the pipeline" toward science careers has been noted (Seymour and Hewett, 1998; Seymour, 2001; NRC, 2007).

Although the proposed improvements noted above differ in detail, a remarkably consistent theme is the call to bring student-centered instructional strategies, such as active- and inquiry-oriented learning, into the classroom. Allen and Tanner (2005) define active learning as "seeking new information, organizing it in a way that is meaningful, and having the chance to explain it to others." This form of instruction emphasizes interactions with peers and instructors and involves a cycle of activity and feedback where students are given consistent opportunities to apply their learning in the classroom. By placing students at the center of instruction, this approach shifts the focus from teaching to learning and promotes a learning environment more amenable to the metacognitive development necessary for students to become independent and critical thinkers (Bransford *et al.*, 2000). A substantial number of studies have shown that active-learning instructional approaches can lead to improved student attitudes (e.g., Marbach-Ad *et al.*, 2001; Prince, 2004; Preszler *et al.*, 2007) and increased learning

DOI: 10.1187/cbe.09–03–0025
Address correspondence to: Peter Armbruster (paa9@georgetown.edu).

outcomes (Ebert-May *et al*., 1997; Hake, 1998; Udovic *et al*., 2002; Knight and Wood, 2005; Freeman *et al*., 2007) relative to a standard lecture format.

The establishment of several national programs that promote active-learning pedagogy (The National Academies Summer Institutes[1] and FIRST II[2]), the establishment of journals such as *CBE—Life Sciences Education*, and the growth of several database repositories of active-learning exercises (MERLOT pedagogy portal[3], TIEE[4], FIRST II, National Digital Science Library,[5] and especially BioSciEdNet[6] and SENCER Digital Libary[7]) are all positive evidence of concerted responses to the calls for change noted above. These resources also provide significant support for faculty committed to implementing active-learning strategies in their courses both in terms of training opportunities and by making example teaching materials readily available. Nevertheless, the proposition of restructuring a large introductory course to emphasize elements of active learning can seem overwhelming for faculty with extensive time commitments in other realms and little or no formal training in pedagogy.

Here, we describe the development and implementation of an instructional design that focused on bringing multiple forms of active-learning and student-centered pedagogies into a traditionally lecture-based introductory biology course. Our course restructuring was motivated by several perceived deficiencies common to traditional lecture-based introductory courses. The most pronounced concern, shared by multiple faculty involved in the course, was poor student attitudes. Both numeric and written responses on course evaluations indicated that students were not satisfied with the course and did not recognize the importance of the course content to their education as biologists. For example, students often commented on course evaluations that the lectures and/or course materials were "boring." Furthermore, individual instructor–student interactions often indicated that students were more concerned with their test scores than with gaining a thorough understanding of the course material. Poor student attitudes also were reflected by poor attendance, limited participation in class, and suboptimal student performance.

We hypothesized that incorporating active-learning and student-centered pedagogy into the instructional design of our course would both improve student attitudes and also lead to increased student performance (Weimer, 2002). We chose to focus primarily on using problem-based learning activities because these activities tend to be more succinct and less open-ended than case-based activities, and thus it was easier to integrate problem-based activities into our previously established lecture organization. Our positive results illustrate how changing the instructional design of a course, without wholesale changes to course content, can lead to improved student attitudes and performance. The goals of this article are to 1) describe the elements of our instructional design that contributed to improved student attitudes and performance; and 2) discuss significant future challenges, so that other educators can learn from our experiences.

MATERIALS AND METHODS

Study Design

The course restructuring we describe pertains to the lecture portion of Introductory Biology II, a one-semester course that typically enrolls between 170 and 190 students (details are provided below). The course was taught in a standard lecture format in 2006 and redesigned to emphasize active learning and student-centered pedagogy in 2007 and 2008. The first author taught the course in all 3 years (2006–2008) and in the 2 years before 2006. The hypotheses we consider in this study are that student attitudes and performance increased in 2007 and 2008 in response to the instructional design we implemented.

Course Description

Introductory Biology II is the second semester of a 1-yr sequence required for biology majors and premedical students. The first semester of the sequence, Introductory Biology I, focuses on molecular and cellular biology with some treatment of development and physiology. Introductory Biology II emphasizes principles of ecology, evolution, and a survey of the diversity of life. This basic course content was not changed substantially as part of the revision we describe, although we modified the order in which the material was presented (see below). In all 3 years, the lectures consisted of three 70-min periods per week. There also was an optional weekly recitation section where the instructor was available to answer student questions. In all 3 years (2006–2008), we handed out a set of questions ("the daily dozen") for each lecture to help guide students in their assigned textbook reading, and discussion in recitation often centered on these questions. Before our course revision, assessment for the lecture portion of the course consisted of three midterms and a final examination, with each exam consisting of a mix of quantitative problem solving, short answer, and short essay questions. As part of our course revision, we modified this assessment plan to include 10 weekly quizzes, two midterms, and a final exam. In all 3 years, all students were required to enroll in a weekly 3-h laboratory section that was assessed and evaluated separately from the lecture portion of the course. The laboratory portion of the course was not a part of this course revision.

Course Redesign

Our course redesign consisted of three major elements:

1. **Reorder course content.** We reordered the presentation of the course content in an attempt to teach specific content within the context of broad conceptual themes. For example, a new lecture on evolutionary developmental biology ("evo-devo") was presented before the series of lectures surveying animal diversity. This lecture was designed to both serve as an intellectual bridge between the sections of the course describing evolutionary mechanisms and organismal diversity and also to help students understand patterns of animal diversity by understanding some of the mechanisms by which that diversity evolved. As another example, two lectures on photosynthesis were presented immediately following a lecture on ecosystem ecology in order to help students understand the details and importance of primary productivity within the context of nutrient cycling of ecosystems. We also ended the course with a two-lecture module on the biology of avian flu that synthesized a number of topics taught during

---

[1] http://dels.nas.edu/summerinst/index.shtml.
[2] http://first2.plantbiology.msu.edu//.
[3] http://pedagogy.merlot.org/.
[4] http://tiee.ecoed.net/.
[5] http://nsdl.org/.
[6] www.biosciednet.org/portal/.
[7] www.sencer.net/search.cfm

different parts of the semester. In these lectures we emphasized the role of mutation and reassortment in viral evolution, discussed how species interactions influenced viral reassortment, and considered epidemiological models of viral transmission and spread. A copy of the course syllabus is available by request from the corresponding author.

2. **Active learning and group problem solving.** We incorporated active and problem-based learning into every lecture. Students were organized into groups of four on the first day of class, asked to sit together throughout the semester, and in almost every lecture groups were presented with a quantitative or conceptual problem. Examples of a quantitative problem concerning Hardy–Weinberg equilibrium and a strip sequence problem (Handelsman *et al*., 2007) concerning character displacement are presented in Table 1, A and B, respectively. Group problems were typically displayed on a PowerPoint slide, and the groups were given 3–5 min to work on the problem. During this period, the instructor would move from group to group in the classroom to monitor student progress and offer suggestions if a group encountered difficulty. The level of student activity was clearly indicated by the noise level of student discussions in the classroom, which was monitored to determine when to bring the group work to a close. Haphazardly selected group representatives were then asked to report out to the class after each group problem-solving session. In addition to the examples in Table 1, we used a variety of active-learning exercises as described in Handelsman *et al*. (2007), including think-pair-share, 1-min papers, and concept maps.

   A personal response system (a.k.a. "clickers") also was used to promote active learning in the classroom. Each lecture included two to six "clicker questions" that were presented as multiple-choice questions on a PowerPoint slide. Generally, we developed the questions to address a specific concept covered in the lecture, but in some cases Graduate Record Exam (GRE) or Medical College Admission Test questions were presented with a label indicating the source of the question. Effective implementation of clicker questions is discussed below (see *Discussion*), and two representative examples are presented in Table 2. Students were awarded participation points (20 points of a course total of 700 points) if they answered $\geq$75% of all clicker questions presented over the entire semester (approximately 120 total questions each year), regardless of whether their answers were correct. Clickers were also used to administer weekly quizzes (see below).

3. **Student-centered pedagogy.** We adopted several additional strategies to create a more student-centered learning environment. Every lecture included a set of learning goals made explicit to students in the lecture PowerPoint slides (Table 3). All exam and quiz questions were then labeled with the corresponding learning goals to emphasize the alignment of learning goals and assessment. We also included a set of vocabulary terms for each lecture to help the students focus on important concepts, and with the hope that students would use these technical terms to formulate more precise and succinct answers to free-response questions on exams. We also placed an increased emphasis on formative assessment by integrating assessment and self-assessment components into activities during lecture so that students would receive feedback designed to improve their performance (Handelsman *et al*., 2007). For example, virtually every group problem in lecture (e.g., Table 1) included a component of formative assessment because we always discussed the answer to each problem in class and the group work problems closely resembled the problems on exams. Finally, we administered 11 weekly quizzes worth 8 points each, with only the top 10 scores applied to the final grade. These weekly quizzes thus provided regular feedback on student performance in a "low-stakes" assessment environment and encouraged students to keep up with the material on a regular basis.

The course redesign was implemented for the first time in 2007. In 2008, the course organization closely followed that of 2007, with minor modifications based on student feedback in 2007 (see *Discussion* for details). The most substantial change in 2008 involved moving the weekly quizzes to Thursday, the day after the optional recitation session (in 2007 quizzes were administered on Tuesday).

Table 1. Examples of problems administered to groups during lecture

A. A three-allele problem on Hardy–Weinberg equilibrium

Problem: You are studying a population of wildflowers with three alleles (A, a, A′) where p = freq (A), q = freq (a), r = freq (A′).
AA genotypes are blue, aa genotypes are red, A′A′ individuals are pink, and all heterozygous genotypes are purple.
You sample 100 flowers and find four blue and 16 red flowers.
If this population is in Hardy–Weinberg-equilibrium, what is the frequency of the AA′ heterozygote genotype?

Answer: The frequency of AA′ genotype is 0.16.

B. A modified strip-sequence problem concerning character displacement

Problem: Consider two species of birds that occupy a similar ecological niche. Furthermore, these birds occur both in sympatry and in allopatry. Arrange the following terms in the correct order to describe the processes leading to character displacement (note: one term does not fit).

character displacement—competition—competitive exclusion—resource partitioning—sympatric populations—natural selection—evolutionary change

Answer: Sympatric populations → competition → natural selection → evolutionary change → resource partitioning → character displacement

For both problems, students were asked to work collaboratively in preassigned groups for 5 to 10 min to solve the problem. A single successful group was then asked to describe their solution to the class. The instructor then asked for explanations from any other group that successfully solved the problem by using an alternative approach or reasoning.

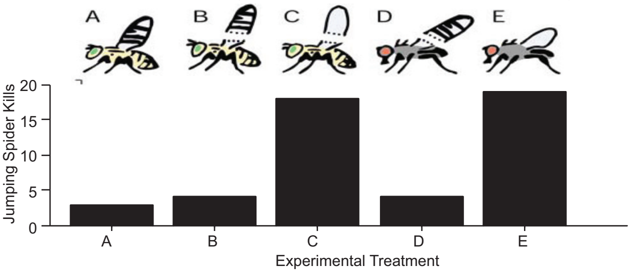
Assessment of Student Attitudes and Performance

We assessed student attitudes toward the course in all 3 years by 1) administering a three-page questionnaire that used both Likert-scale and free response questions (see Supplemental Material I), and 2) comparing scores on university-administered course evaluations for questions that addressed student satisfaction. We assessed student performance by comparing class scores on two identical final exam questions administered in 2006, 2007, and 2008 (see Supplemental Material II). We do not hand back the approximately 20-page final exam, so it is not possible that the specifics of these questions were available to students in later years (i.e., 2007 and 2008). We chose final exam questions that addressed the fundamental topics of logistic population growth and life-history trade-offs and a more conceptual question on island biogeography (see Supplemental Material II). All of these topics were emphasized heavily in lecture in all 3 years. In addition, two educational experts scored all of the questions on the final exam in 2006, 2007, and 2008 according to Bloom's taxonomy of learning (Bloom, 1956). Bloom's taxonomy identifies six hierarchical levels of understanding that range from knowledge (level 1), to comprehension (level 2), application (level 3), analysis (level 4), synthesis (level 5), and evaluation (level 6). We used a weighted Kappa statistic (Altman, 1991) to quantify the interrater reliability because this statistic is appropriate for ratings that fall into discrete categories. Because the original scored exams are no longer available, direct comparison of performance on questions that differ in Bloom's ranking across years is not possible.

Table 2. Examples of clicker questions administered during lecture

A. A question administered during a lecture on adaptation. See pp. 368–369 in Freeman and Herron (2007) for further details.

Recall the experiment examining predation of the tephritid fly by jumping spiders. Now, consider the data below which differ from the results discussed in lecture. Which of the following conclusions is BEST supported by the data below?



A. Behavior but not wing pattern reduces kills by jumping spiders.
B. Wing pattern but not behavior reduces kills by jumping spiders.
C. NEITHER wing pattern nor behavior reduce kills by jumping spiders.
D. BOTH wing pattern and behavior are required to reduce kills by jumping spiders.
E. Cutting off and regluing the wing affects the kills by jumping spiders.

B. A question administered during a lecture on photosynthesis.

C4 and CAM photosynthesis are similar in that

A. In both C4 and CAM the light reactions produce ATP and NADPH to drive the Calvin cycle during the daylight hours.
B. In both C4 and CAM, the enzyme PEP carboxylase fixes CO_2 into a four-carbon organic acid.
C. In both C4 and CAM, the Calvin cycle is most active during the daylight hours.
D. B and C only.
E. All of the above.

Additional examples are discussed in *Discussion*.

However, our electronic grade book does permit us to compare the proportion of points at different Bloom's levels and performance on the final exam in all 3 years.

Data Analysis

We tested for differences in class composition between years based on categories in Table 4 by using a χ^2 goodness-of-fit test. We tested for differences in Likert-scale student responses concerning attitudes toward the course from both the questionnaire and university course evaluations by using one-way analysis of variance (ANOVA) followed by a posteriori comparison of means with a sequential Bonferroni correction to control for experiment-wise error ($\alpha = 0.05$). We used a one-way ANOVA of Likert-scale ratings of the helpfulness of different lecture components (i.e., weekly quizzes, clickers, etc.) with lecture component as a fixed effect and students nested within lecture component as a random effect. The one-way ANOVA was followed by planned (a priori) comparisons of means of different lecture components (averaged across years) with Bonferroni correction for multiple comparisons. We tested for differences between years (2007 and 2008) for each individual lecture component with a Student's t test, again with Bonferroni correction for multiple comparisons. To test for differences in student performance on identical final exam questions among years (2006, 2007, and 2008), we also used one-way ANOVA and a posteriori comparison of means with Bonferroni correction. To test for differences in performance on the entire final exam in 2006, 2007, and 2008, we performed one-way ANOVA on square-root arcsine-transformed percentage scores in each year followed by a posteriori comparison of means with Bonferroni correction.

Students' free responses on the questionnaire (see Supplemental Material I) provided a source of qualitative data on attitudes. Each student's answers to question 5A ("What specifically did you like about the course?") and positive comments from question 6 ("What else would you like to tell us?") were combined to reflect the student's positive feedback. Similarly, each student's answers to question 5B ("What specifically did you dislike about the course?") and negative comments from question 6 were combined to represent the student's negative feedback. Negative and positive feedback for each of the 3 years was coded separately. Codes were developed in vivo (Strauss and Corbin, 1990). Most codes reflected specific course or lecture elements, such as clickers, quizzes, learning goals, PowerPoint slides, and guest lectures. Additional codes were developed to tag students' more general or descriptive statements such as: "too early" or "connected to the real world." The open response nature of these questions meant that individual statements could be tagged with several codes (examples of coded text are provided in Supplemental Material III).

Categories were developed in relation to code frequencies as determined by the number of students whose statements were tagged with that code in a given year. In this way, we avoided overestimating a code's frequency when, for example, an individual mentioned a lecture element multiple times. The most frequently used codes (e.g., clickers and quizzes) were elevated to category status. Text tagged with these codes were re-examined for the explanatory details (subcodes) that are presented associated with each category in Tables 5 and 6.

RESULTS

Composition of Student Body

Approximately 60% of Introductory Biology II students are in the first year of undergraduate study; 75% identified themselves as premedical students, and only 40% were declared biology majors. The student composition of the course (Table 4) in 2007 and 2008 did not differ significantly from 2006 ($\chi^2 = 11.21$, df = 7, $p > 0.10$).

Table 3. Examples of learning goals presented at the beginning of each lecture

A. Four representative learning goals (of eight total) for a lecture on random genetic drift:

1. Be able to explain the details of the Buri (1956) experiment and the major results (including interpretation of graphs) and conclusions. Also, be able to make predictions about how the results of the experiment would have been different if certain aspects of the experiment were modified.
2. Be able to formulate basic predictions about how allele frequencies will change within or between populations based on knowledge of population size(s) and/or the action genetic drift.
3. Be able to explain the concept of inbreeding depression and the genetic basis of inbreeding depression.
4. Be able to explain the example of inbreeding depression in the Florida panther discussed in class. What are the proximate and ultimate causes of inbreeding depression in this example? What are the potential management strategies to address the deleterious effects of inbreeding depression in this example?

B. Four representative learning goals (of eight total) for a lecture on chordate evolution and diversity:

1. Be able to identify and explain the distinguishing characteristics of the phylum Chordata.
2. Be able to draw/explain the chordate phylogeny constructed in class, including accurate placement of taxa and evolutionary innovations along the tree.
3. Be able to explain the functional significance of evolutionary innovations from 2 above.
4. Be able to describe the aspects of amphibian anatomy, physiology, and life history that keep them restricted to moist habitats.

Student Attitudes

All measures of student satisfaction differed significantly between years (Figure 1). These measures include change in interest in the course material from the start to the end of the semester ($F_{2409} = 5.22$, $p < 0.001$), ranking of relevance of course material to long-term student goals ($F_{2407} = 6.65$, $p = 0.001$), self-reported student learning ($F_{2355} = 11.70$, $p < 0.001$), ranking of classroom presentations as stimulating ($F_{2358} = 26.52$, $p < 0.001$), ranking of the course as challenging ($F_{2355} = 15.87$, $p < 0.001$), and overall evaluation of instructor ($F_{2355} = 15.87$, $p < 0.001$). For all measures of student satisfaction, a posteriori comparison of treatment means indicated that student satisfaction was significantly higher in 2007 and 2008 than in 2006 (sequential Bonferroni, $p < 0.05$) but did not differ between 2007 and 2008 ($p > 0.05$; Figure 1). A summary of student free responses to questions probing student satisfaction and dissatisfaction in all 3 years is provided in Tables 5 and 6, respectively.

Table 4. Class composition of Introductory Biology II in 2006, 2007, and 2008[a]

	2006	2007	2008
Freshman, %	63	62	58
Biology majors, %	36	36	43
Other science majors, %	26	21	15
Premedical students, %	76	74	75
Total student responses[b]	122	133	153

[a] Based on student responses to the questionnaire in Supplemental Material I.

[b] Total enrollment was 165 in 2006, 179 in 2007, and 176 in 2008.

Student-centered and Active-Learning Components

Student ranking of the helpfulness of different lecture components (Figure 2) indicated significant differences among components ($F_{7,2245} = 129.64$; $p < 0.001$). Planned (a priori) comparisons of ranking scores indicated that across both years, learning goals were considered the most helpful lecture element, followed by clicker questions and weekly quizzes, which did not differ significantly. The vocabulary list and "daily dozen" reading questions were ranked least helpful and did not differ significantly. Group work, recitation, and outside class study groups received intermediate rankings. Planned comparisons of specific lecture elements between years indicated significant differences in the helpfulness ranking between 2007 and 2008 for the vocabulary list ($t = 3.19$, df = 284, $p < 0.01$) and recitation ($t = 6.06$, df = 284, $p < 0.001$). All other lecture components did not differ in helpfulness ranking between 2007 and 2008.

Student Performance

Student performance on identical final exam questions (e.g., see Supplemental Material II) was greater in years when the material was taught in an interactive format (2007 and 2008) than in 2006 when the material was taught in a standard lecture format (Figure 3). Student scores on questions concerning logistic population growth and life-history tradeoffs differed significantly among years ($F_{2506} = 36.97$, $p < 0.001$) and were higher in 2007 and 2008 than in 2006 ($p < 0.05$; Figure 3A). Student scores differed significantly among years on a question concerning island biogeography ($F_{2505} = 14.55$, $p < 0.001$), with scores in 2008 higher than scores in 2006 and 2007, which did not differ significantly ($p > 0.05$; Figure 3B).

The Bloom's taxonomy scores assigned to final exam questions by two independent raters yielded moderate (Altman, 1991) interrater reliability (weighted Kappa = 0.54), with the majority of disagreements (86%) due to differences between ratings at Bloom's levels 1 and 2. Therefore, for each final exam, we pooled the number of points available across lower-level (1–2, knowledge-comprehension) and higher-level (3–4, application-analysis) Bloom's categories. In 2006 and 2007, 82–85% of the final exam points consisted of lower-level Bloom's categories and 15–18% were higher-level Bloom's categories. In 2008, 75% of the final exam points consisted of lower-level Bloom's categories and 25% were higher-level Bloom's categories. Student performance on the final exam differed significantly among years ($F_{2442} = 12.24$, $p < 0.001$). Despite the higher proportion of points associated with higher-level Bloom's categories in 2008, performance in 2008 (average score = 91%) was significantly higher ($p < 0.05$) than in 2006 (86%) and 2007 (85%), which did not differ ($p > 0.05$).

Table 5. Top five categories of positive free response (and associated explanatory subcodes) regarding student satisfaction from questions 5A and 6 on the questionnaire provided in Supplemental Material I

2006 (n = 107)		2007 (n = 114)		2008 (n = 148)	
56%	Traditional course materials: Power Point, videos, handouts, outlines	24%	Quality of instruction: enthusiasm, clarity, organization, comprehensive, pacing	27%	Quality of instruction: organization, clarity, and interesting
27%	Quality of instruction: enthusiasm, clarity, and organization	14%	Clickers: clarified, engaged, immediate feedback	16%	Clickers: engaged, feedback
8%	Overall well done	13%	Additional course materials: learning goals, groups, quizzes	12%	Guest lectures
6%	Interesting material	12%	Content interesting: range of topics, specific topics	12%	Interaction in lecture: via group activities, clickers, multiple approaches to learning
4%	Specific topics	11%	Traditional course materials: PowerPoint, videos	10%	Traditional course materials: PowerPoint, videos, recitation

See text for details of classification of categories.

DISCUSSION

A traditional lecture format in a large introductory classroom often emphasizes content rather than process and in doing so often fails to convey to students the nature of hypothesis-based inquiry which is at the heart of scientific research. There is reason to believe that this deficit diminishes learning outcomes and may contribute to the loss of some of our most talented students at the introductory level (NRC, 2003, 2007; Handelsman *et al*., 2007). The primary goal of our course restructuring was to improve student attitudes in the course, motivated by the hypothesis that improved attitudes would lead to improved learning outcomes (Weimer, 2002). The course reorganization we described sought to address these challenges by 1) reorganizing the course material to emphasize context, 2) engaging students with active learning in every lecture, and 3) creating a more student-centered classroom environment.

Student Attitudes

The data in Figure 1 clearly indicate that the changes we implemented in 2007 and 2008 improved student attitudes toward the course. For every question considered, student satisfaction scores increased significantly between 2006 and 2007 and did not differ between 2007 and 2008. It is important to note that in 2006, the first author was teaching this course for the third consecutive year. University teaching evaluation scores were consistent in the three years before 2007, and in fact a major reason for implementing the changes we describe in 2007 was that the instructor (first author) felt strongly that after 3 years, additional teaching experience alone was unlikely to cause a significant change in student response to the course. We therefore attribute the clear and consistent changes in student attitudes between 2006 and 2007 (Figure 1) directly to the elements of course redesign we describe here, and the similarity of student responses in 2007 and 2008 (Figure 1 and Tables 5 and 6) further supports this interpretation.

The students' free-responses summarized in Tables 5 and 6 are consistent with the data presented in Figure 1. First, it should be noted that the proportion of positive comments increased from 2006 (65%) to 2007 (81%) and 2008 (89%). In 2006, the top category (56%) of positive response concerned traditional course material (e.g., PowerPoint slides, videos), whereas in 2007 and 2008 traditional course materials were mentioned in only 10–11% of the positive comments, and quality of instruction was the most common positive com-

Table 6. Top five categories of negative free response (and associated explanatory subcodes) regarding student satisfaction from questions 5B and 6 on the questionnaire provided in Supplemental Material I

2006 (n = 104)		2007 (n = 101)		2008 (n = 118)	
25%	Lecture not stimulating	22%	Group work	17%	Group work
17%	Exams: too hard, too specific, grading	15%	Quizzes: too frequent, too hard, format	15%	Quizzes: points, stressful, too hard
13%	Logistics: too early, too long, use board	13%	Logistics: too early, too long, no breaks	11%	Logistics: lecture too long
12%	Course materials: improve handouts/outlines/PowerPoint	11%	Guest lectures: irrelevant, unrelated to course material	10%	Difficult to know what to study, learning goals incomplete
6%	Subject matter not interesting	11%	Course materials: PowerPoint, folders, use the board, more movies and articles	8%	Spent too much time on easy topics, not enough time on hard topics

See text for details of classification of categories.

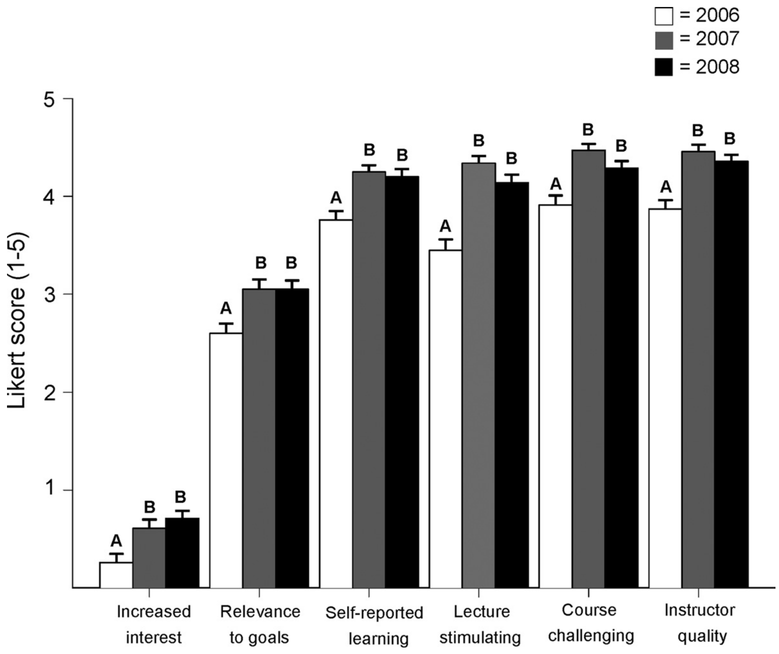


Figure 1. Mean ± SE student-reported attitudes from 2006 (standard lecture format), 2007, and 2008 (revised lecture format). "Increased interest" and "Relevance to goals" were questions on an instructor administered questionnaire (see Supplemental Material I), all other questions were part of the university course evaluation (see text). Increased interest represents difference in interest in the subject matter after taking the course relative to interest before taking the course (i.e., questions 2A and B, see Supplemental Material I). For each question, comparison among years (2006, 2007, and 2008) was significant ($p < 0.001$) by one-way ANOVA. Results of a posteriori comparison of means for each question indicated by letters where means that share the same letter are not significantly different ($p > 0.05$) with Bonferonni correction for experiment-wise error ($\alpha = 0.05$).

ment in both years at 24 and 27%, respectively (Table 5). Together, these results clearly indicate that students' perception of the quality of instruction increased in 2007 and 2008, similar to results in Figure 1.

Student-centered and Active-Learning Components

Students' positive free-response answers explicitly referencing specific components of the course redesign were the second (14%) and third (13%) most frequent category of positive response in 2007, and second (16%) and fourth (12%) most frequent category of positive response in 2008 (Table 5). These comments in 2007 and 2008 that specifically mention the active-learning and student-centered pedagogy we introduced in 2007 included references to "engagement," "immediate feedback," and "multiple approaches to learning." There were almost none of these specific references in 2006.

With respect to negative free responses (Table 6), in 2006 the most frequent category of response was that lecture was not stimulating (25%), whereas in 2007 and 2008 that category composed <1% of the negative responses. Again, these data corroborate the results in Figure 1, where lecture was ranked as more stimulating in 2007 and 2008 than in 2006.

It is important to note, however, that two specific elements of our course redesign were explicitly mentioned as the first (group work, 17%) and second (weekly quizzes, 15%) most frequent category of negative response in both 2007 and 2008. Group work was also ranked relatively low in terms of helpfulness to student learning (Figure 2). Our interpretation of the feedback on group work is that we need to further refine this element of the course. We adopted strategies for effectively implementing group work as discussed by Handelsman *et al.* (2007) and Ebert-May and Hodder (2008). Students did not receive credit for these in-class active-learning exercises, but the requirement to report out to the class seemed to provide a strong incentive for most students to engage seriously in these activities. During each group-work exercise the instructor would move throughout the classroom to monitor group progress, and it was rare to find a group that was not seriously engaged in the exercise. However, the attempt to include a group exercise in almost every lecture meant that both the quality and rigor of exercises varied considerably. The group exercises that elicited the most animated student participation were those that were sufficiently challenging that very few students could solve the problem individually, but at least 50% or more of the groups could solve the problem by working as a team. Some of our most active group interactions occurred when we administered a challenging quiz, and then immediately allowed the students to retake the quiz as a group with the stipulation that the students would receive the highest of either their group or individual scores. This consideration suggests that a potential modification to further increase engagement in the group work would be to assign points to these in-class exercises (Ebert-May and Hodder, 2008).

Our interpretation of the relatively high proportion of negative comments regarding the weekly quizzes (Table 6) differs from that regarding the group work. The weekly quizzes were implemented in order to encourage students to keep up with the course material and to provide them with regular feedback on their understanding of the material in a low-stakes assessment environment. Note that in Figure 2 students ranked the weekly quizzes third highest in terms of helpfulness in both 2007 and 2008. We thus interpret the data in Table 6 and Figure 2 to indicate that although some students may dislike the weekly quizzes (administered at 8:50 AM), many recognized that these quizzes were helpful to their learning. Sixty-four percent of the respondents rated

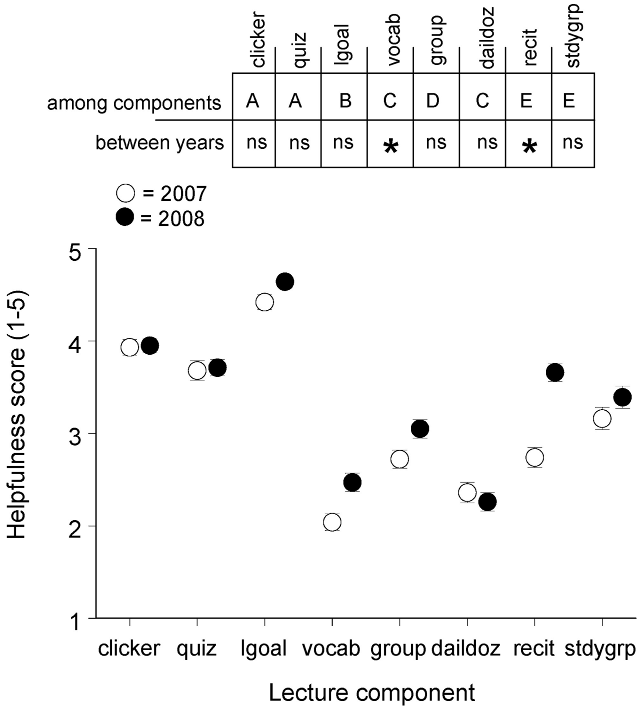


Figure 2. Mean (± SE) ranking of helpfulness for lecture components by students in 2007 (○) and 2008 (●). Lecture components as described in text: clicker = clicker questions; quiz = weekly quizzes; lgoal = learning goals; vocab = vocabulary lists; group = group work in class; daildoz = daily dozen reading questions; recit = optional weekly recitation section; and stdygrp = optional study group outside class. Results of one-way ANOVA testing for differences among lecture components was highly significant ($p < 0.001$). The legend on the top indicates results of planned comparisons (A) among components (pooled across years) where components that share the same letter are not significantly different ($p > 0.05$) and (B) between years (2007 and 2008) for each individual component, where * indicates $p < 0.05$, ns indicates $p > 0.05$.

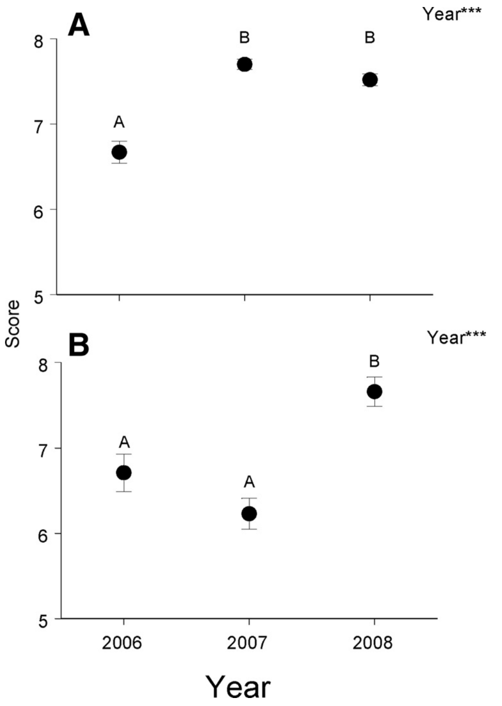


Figure 3. Mean (± SE) points scored on identical final exam questions administered in 2006, 2007, and 2008. (A) Logistic growth and life-history evolution, (8 points possible). (B) Island biogeography, see Supplemental Material II (9 points possible). The legend in the top right indicates the results of one-way ANOVA testing for differences among years, ***$p < 0.001$. Results of a posteriori comparison among years indicated by letters, means associated with the same letter are not significantly different ($p > 0.05$).

quizzes at 4 or 5 in terms of their helpfulness. The following quote is a typical comment made by students who rated quizzes at 4 or 5:

"Quizzes seemed like a hassle at first but in the end when our exams came up, since I had been studying all along for the quizzes, I had learned/studied most of the material, so I actually appreciate the weekly quiz system." S136.2008.Q3B

We view these results as positive evidence of metacognitive awareness (Bransford *et al.*, 2000) in that the weekly quizzes seem to have helped these students identify strategies for enhancing their own learning. This represents a particularly important goal for introductory classes that aim to prepare students for more advanced course work and independent learning.

Figure 2 indicates considerable consistency between 2007 and 2008 in the ranking of various lecture elements in terms of the helpfulness to student learning. The explicit learning goals (Table 3) ranked highest in both years (Figure 2). From a student's perspective, learning goals establish clear expectations about what skills and content students should master from each lecture. From an instructor's perspective, learning goals play a critical role in shaping both instructional activities and assessment through the process of "backward design" (Wiggins and McTighe, 1998; Handelsman *et al.*, 2007), whereby learning goals explicitly articulate the desired learning outcomes to both instructor and students. Those desired outcomes then specify the assessment tasks that determine whether the desired outcomes have been met, and also shape teaching activities required to meet the desired goals. During 2007 and 2008, through the process of backward design, the learning goals provided a clear "road map" for both determining the content and organization of lectures and also for writing exams, whereas in 2006 both processes took place in a much less structured manner.

The personal response system (clickers) ranked the second highest in terms of helpfulness with learning in both 2007 and 2008 (Figure 2). These results are consistent with those of a large number of previous studies documenting positive student responses to clicker systems (for review, see Judson and Sawada, 2002) and a large body of evidence indicating that the use of clickers and associated peer interaction (see below) can lead to improved student learning (Crouch and Mazur, 2001; Knight and Wood, 2005; Preszler *et al.*, 2007; Smith *et al.*, 2009). In a recent and intriguing study from physics, Reay *et al.* (2008) found that the use of clickers not only led to increased learning gains in an introductory physics course but also seemed to reduce the performance difference between males and females.

The clickers were an effective pedagogical tool in our introductory biology course in several respects. First, the clicker system provided "real-time feedback" to the students (Table 5). This feedback allowed the instructor to establish clear expectations regarding the depth of student understanding required to answer quiz and exam questions correctly. Simultaneously, this information allowed students to gauge their understanding continually relative to those expectations (i.e., formative assessment). The clickers were also extremely helpful in identifying, and thus allowing us to rectify, by addressing in a more direct and thorough manner, student misconceptions. Two striking misconceptions in our class concerned the ability to interpret a phylogenetic tree (see the "tree thinking" exercise by Baum *et al*., 2005) and the failure to recognize that photosynthetic organisms not only fix CO_2 through photosynthesis but also release CO_2 through cellular respiration (Wilson *et al*., 2006).

The clickers were also very useful in initiating peer instruction in the classroom (Mazur, 1997; Crouch and Mazur, 2001). This occurred when between 35 and 75% of the class answered a clicker question incorrectly, and students were then instructed to consult with a neighbor for 1 to 2 min to discuss their answers. The students were then repolled without being informed of the correct answers. Such occasions invariably led to animated discussion among the students in the class, and almost always resulted in an increase in the proportion of correct answers when the students were repolled. The clicker questions that generated the most animated student discussion were those that either did not have a single correct answer or that elicited a relatively even number of responses between two or more answers. These results are consistent with previous studies that demonstrate the efficacy of peer instruction facilitated by clickers to promote student learning (Crouch and Mazur, 2001; Freeman *et al*., 2007; Smith *et al*., 2009). However, it is critical to recognize that it is the peer interaction rather than the clickers per se that promotes student learning (Smith *et al*., 2009), emphasizing that an appropriate underlying pedagogical design is essential for the effective use of clickers (Mazur, 1997; Crouch and Mazur, 2001).

Student ranking of helpfulness for the vocabulary list and recitation increased significantly from 2007 to 2008 (Figure 2). Notably, these components were ranked relatively low in 2007 and this feedback from the student questionnaire in 2007 enabled us to target these aspects of the course design in 2008. The vocabulary list presented at the beginning of each class (see Course Redesign) was ranked as the least helpful element of the lecture in 2007. The goal of these lists was to help students use technical terminology to formulate concise and precise answers to free-response questions on exams. In 2008 we discussed this goal in lecture and explicitly modeled the process several times. In the future, we plan to develop active-learning exercises that explicitly focus on clear written communication.

The increase in helpfulness ranking from 2007 to 2008 for the optional recitation session (Figure 2) was particularly notable. Based on student feedback in 2007, in 2008 we moved the quizzes to Thursday so that they were given the day immediately after the recitation sessions. Attendance at the recitation sessions increased dramatically, consistent with the data in Figure 2. This change from 2007 to 2008 provides an excellent example of how student feedback can be used to make simple changes that have a large impact on student satisfaction and performance.

Finally, we note that one of the major elements from the course in 2006 that was carried over into 2007 and 2008 was the daily dozen (a list of questions designed to help students identify important concepts in the textbook reading assignments), which ranked relatively low in terms of helpfulness compared with elements that were introduced as part of our course restructuring in 2007. We did not receive specific positive or negative feedback regarding the daily dozen on our questionnaire in 2006–2008 (Tables 5 and 6), and we attribute the relatively low ranking of this component in 2007 and 2008 (Figure 2) to greater enthusiasm for other components of the course.

Student Performance

Our data on academic performance are consistent with previous studies indicating that student-centered pedagogy and interactive-learning activities increase student performance (Ebert-May *et al*., 1997; Udovic *et al*., 2002; Knight and Wood, 2005; Freeman *et al*., 2007; Walker *et al*., 2008). The data in Figure 3 illustrate student performance on identical final exam questions administered in all 3 years and show consistent increases in performance between 2006 and 2008. Furthermore, the proportion of points on the final exam for questions at higher levels of Bloom's taxonomy (levels 3–4, application-analysis) increased from 15–18% in 2006–2007 to 25% in 2008. Furthermore, the average student performance on the final exam also increased in 2008 (91%) relative to 2006 (86%) and 2007 (85%). Together, these results indicate increased academic performance and imply increased proficiency with higher-order problem-solving skills associated with the changes in instructional design implemented in our course. These conclusions are somewhat conservative because the 2006 final exam contained a section in which students were allowed to choose six of eight questions to answer, but students were not given any choices on the 2007 and 2008 final exams.

The results on student performance noted above suggest that the most pronounced increases in performance occurred between 2007 and 2008, whereas results in Figure 1 and Table 5 indicate that student attitudes increased significantly from 2006 to 2007 and did not change between 2007 and 2008. We believe these results indicate that a semester of experience with implementing the active-learning and student-centered pedagogies in 2007 made these approaches more effective in improving student performance in 2008. Although the initial goal of our course redesign was to target student attitudes, we are now initiating more intensive efforts to quantify student learning by using pre- and postcourse assessment tools, assessment of higher-order skills such as the interpretation of primary literature, and performance on the Biology GRE.

Institutional Context

The course redesign we implemented required a significant time investment both in the approximately 6 mo leading up to 2007, and during the first semester of implementation. Attendance at a national workshop, the National Academies Summer Institutes on Undergraduate Education in Biology (www.academiessummerinstitute.org/) provided significant background theory and training. Also, in fall 2006, we convened a series of on-campus seminars featuring national leaders in science education (http://cndls.georgetown.edu/events/symposia/TFU/). These seminars were particularly useful both in generating support from our departmental colleagues to implement changes in a course that is foundational to the department's curriculum and also in providing the opportunity to discuss specific details of course redesign with individuals highly experienced in implementing active-learning and student-centered pedagogical approaches. It is important to note, however, that once the initial course redesign was implemented in 2007, teaching the course in 2008 did not require a significant additional time commitment relative to 2006 (before our changes were implemented) and yet the increased positive student response to the course was sustained (Figure 1 and Table 5). Furthermore, the improved scores on the university-administered course evaluations (see questions 2–6 in Figure 1), the primary mechanisms of assessing teaching at most institutions, indicates that the time investment required to implement a course restructuring can have a positive impact on instructor evaluation criteria.

Finally, the course redesign had another unanticipated benefit: it improved not only the students' attitude toward the course but also the instructor's morale and enthusiasm. Introductory Biology II has long been a problematic course for our department because of deficiencies noted in the *Introduction* (poor student attitudes, passive [superficial] learning, and suboptimal student performance). As a consequence, instructors often lose enthusiasm for teaching this course after 2 to 3 years. However, the interactive pedagogy and positive student responses made this a much more exciting and rewarding course to teach in 2007 and 2008.

The changes we implemented also have had an impact at the departmental level. Based in part on the positive student reactions to interactive and student-centered pedagogy in Introductory Biology II, four instructors have implemented the use of clickers in their courses and one faculty member attended the 2007 National Academies Summer Institutes on Undergraduate Education in Biology.

In summary, we developed and implemented an instructional design that focused on incorporating active-learning and student-centered pedagogy into what was previously a traditional lecture-based introductory biology course. These changes led to sustainable improvements in student attitudes and performance. Although the changes we implemented required a significant time commitment in the first year (2007), this was essentially a "one time investment" because it did not require extra effort to teach the course using the revised model in 2008. Furthermore, several faculty in our department have begun to incorporate interactive and student-centered pedagogies into their courses. The course reorganization we describe thus not only provides a model for revision of an individual course but can also provide a catalyst for institutional reform.

ACKNOWLEDGMENTS

We thank Janet Russell and Michael Hickey for scoring our final exam questions according to Bloom's taxonomy; and Diane Ebert-May, Jay Labov, Janet Russell, Debra Tomanek, and two anonymous reviewers for helpful comments on previous versions of this manuscript. We also acknowledge Diane Ebert-May, Diane O'Dowd, and Robin Wright for valuable input on our course restructuring. Financial support was provided by the College Curriculum Renewal Project of Georgetown College and the Undergraduate Learning Initiative from the Center for New Designs in Learning and Scholarship, Georgetown University.

REFERENCES

Allen, D., and Tanner, K. (2005). Infusing active learning into the large-enrollment biology class: seven strategies, from the simple to complex. Cell Biol. Educ. *4*, 262–268.

Altman, D. G. (1991). Practical Statistics for Medical Research, London, United Kingdom: Chapman & Hall.

American Association for the Advancement of Science (1989). Science For All Americans: A Project 2061 Report on Literacy Goals in Science, Mathematics and Technology. Washington, DC.

Baum, D. A., Smith, S. D., and Donovan, S.S.S. (2005). The tree-thinking challenge. Science *310*, 979–980.

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

Boyer, E. L. (1998). The Boyer Commission on Educating Undergraduates in the Research University, Reinventing Undergraduate Education: A Blueprint for America's Research Universities, New York: Stony Brook.

Bransford, J. D., Brown, A. L., and Cocking, R. R. (2000). How People Learn: Brain, Mind, Experience, and School. Committee on Developments in the Science of Learning. Washington, DC: National Academies Press.

Buri, P. (1956). Gene frequency in small populations of mutant *Drosophila*. Evolution *10*, 367–402.

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

Ebert-May, D., Brewer, C., and Sylvester, A. (1997). Innovation in large lectures teaching for active learning. Bioscience *47*, 601–607.

Ebert-May, D., and Hodder, J. (eds.) (2008). Pathways to Scientific Teaching, Sunderland, MA: Sinauer Associates.

Freeman, S., and Herron, J. C. (2007). Evolutionary Analysis, Boston, MA: Pearson-Benjamin Cummings.

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

Handelsman, J., *et al*. (2004). Scientific teaching. Science *304*, 521–522.

Handelsman, J., Miller, S., and Pfund, C. (2007). Scientific Teaching, New York: W.H. Freeman.

Hake, 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.

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

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

Marbach-Ad, G., Seal, O., and Sokolove, P. (2001). Student attitudes and recommendations on active learning: a student-led survey gauging course effectiveness. J. Coll. Sci. Teach. *30*, 434–438.

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

National Research Council (NRC) (1999). Transforming Undergraduate Education in Science, Mathematics, Engineering and Technology, Washington, DC: National Academies Press.

NCR (2003). Bio 2010, Transforming Undergraduate Education for Future Research Biologists. Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century, Washington, DC: National Academies Press.

NRC (2007). Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future. Committee on Prospering in the Global Economy of the 21st Century: An Agenda for American Science and Technology, Washington, DC: National Academies Press.

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.

Prince, M. (2004). Does active learning work? A review of the research. J. Eng. Educ. *93*, 223–231.

Project Kaleidoscope (2006). Recommendations for Urgent Action in Support of Undergraduate Science, Technology, Engineering, and Mathematics. Washington, DC: Project Kaleidoscope.

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

Seymour, E. (2001). Tracking the processes of change in U.S. undergraduate education in science, mathematics, engineering, and technology. Sci. Educ. *86*, 79–105.

Seymour, E., and Hewett, N. M. (1998). Talking About Leaving: Why Undergraduates Leave the Sciences, Boulder, CO: Westview Press.

Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. L., Guild, N., and Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. Science *323*, 122–124.

Strauss, A. L., and Corbin, J. (1990). Basics of Qualitative Research: Grounded Theory Procedures and Techniques, Newbury Park, CA: Sage Publications.

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

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.

Weimer, M. (2002). Learner-Centered Teaching: Five Key Changes to Practice, San Francisco, CA: Jossey-Bass.

Wiggins, G., and McTighe, J. (1998). Understanding by Design. Alexandria, VA: Association for Supervision and Curriculum Development.

Wilson, C. D., Anderson, C. W., Heidemman, M., Merrill, J. E., Merritt, B. W., Richmond, G., Sibley, D. F., and Parker, J. M. (2006). Assessing students' ability to trace matter in dynamic systems in cell biology. CBE Life Sci. Educ. *5*, 323–331.

Wright, R., and Boggs, J. (2002). Learning cell biology as a team: a project-based approach to upper-division cell biology. Cell Biol. Educ. *1*, 145–153.