Assessment of the Impact of Case Studies on Student Learning Gains in an Introductory Biology Course

By Susan Chaplin

Student performance in lecture-based versus case study–based instruction was compared in this study. Case-based teaching that emphasized problem solving and discussion significantly improved student performance on exams throughout the semester and enhanced students' abilities to correctly answer application- and analysis-type questions.

A growing body of evidence indicates that inductive teaching and learning methods (such as case studies, problems, guided instruction, discovery, projects, or presentations by groups) produce more positive learning gains in science classes than traditional lecture methods (see review by Prince and Felder 2006). Classroom methods that promote active-learning strategies give students opportunities to connect new information to their own experience, provide students with models for applying new knowledge, and promote cognitive skills for higher-order thinking (O'Sullivan and Copper 2003; Prince 2004; Allen and Tanner 2005; Knight and Wood 2005; A.C. Smith et al. 2005; Lundeberg and Yadav 2006a, b; Freeman et al. 2007). Case studies, in particular, that use real-life narratives to explore the content of a particular topic area have been a popular means of engaging students in active learning in the classroom. Faculty perceptions of the advantages of using case studies in their science classes are that there are greater gains in critical thinking and learning in their students, as well as greater engagement and enthusiasm for the course content (Yadav et al. 2007). Faculty perceive that their students are learning more content more deeply, are better able to make connections across content areas and apply knowledge to new situations, as judged by performance on exams and in response to open-ended questions and problem solutions (Yadav et al. 2007). Similarly, assessment of students who have been taught by traditional lecture- or case-based teaching has been done by attitudinal surveys, providing information on whether students perceive they have mastered content, been able to demonstrate their understanding, or have increased their skill level (e.g., Student Assessment of Learning Gains survey at *www.wcer.wisc.edu/salgains*.) However, these results are limited because they are just perceptions of learning gains, not empirical measures of them.

How can one prove empirically that case-based teaching leads to better mastery and deeper understanding of material that enables students to transfer knowledge within new contexts? As an experiment in "scientific teaching" (see Handlesman et al. 2004 for definition) to answer this question, I compared several measures of student performance in an introductory biology course on diversity and organismal structure/function as the course changed from a teacher-centered lecture approach to a student-centered approach using case studies as a framework for mini-lectures with class discussion and problem solving for each unit of study.

Methods

Course history and student demographics

I taught the fall semester of the biology majors' core sequence (organismal biology) nine times between 1994 and 2006. This course, as an introduction to the major, was about 75% freshmen, with approximately equal representation of males and females. Lecture sections ranged from 50 to 90 students taught in classrooms or auditoriums, and the content focused on organismal diversity and adaptation organized into four units: evolutionary history of life, sustaining life (metabolism and energetics), surviving challenge (homeostasis, neural, and endocrine regulation), and reproducing life. Based on lessons learned from my previous 18 years of experience teaching all types of college-level biology courses, I taught the introductory biology courses from 1994 to 1999 using a lecture-style format with online quizzes administered once a week; discussion of the quizzes was typically the extent of student participation in the lecture. After using case studies successfully in upper-division biology courses from 1998 to 2002, I adapted case-based instruction to the introductory level when I resumed teaching the introductory course in 2003. The course focus on organismal diversity and adaptation as well as my organization of the material into four units remained the same, as did the student composition of the course (75% freshmen).

I used the resources found on the website of the National Center for Case Study Teaching in Science (*http://ublib.buffalo.edu/libraries/projects/cases/case.html*), and my own case studies to replace portions of the lectures I had previously given, and devoted several days in each of the four units to class discussion of the case studies, promoting a much more interactive environment. For example, instead of lecturing on the content of the text chapter on the Darwinian Revolution, I used a portion of the case study "The Galapagos" (Schiller and Herreid 1999) from the SUNY Buffalo case study collection to explore the concept of evolutionary change in populations by natural and sexual selection. After a brief introduction to the topic (mini-lecture) at the end of one class period, students worked on part of the case study as homework using the text chapter as a reference. We then spent the next class period discussing answers to questions posed in the case study, as well as new information from the chapter. In this way, we covered essentially the same material on which I would have lectured in approximately the same amount of time, but in an active discussion during which students discovered much of the content for themselves.

Materials

In each year of the study, the current edition (4th or 7th) of Campbell and Reece's *Biology* text (1996, 2004) was the primary reference for the course. Exams were given at the end of each unit (comprising 8–10 lecture periods) and were always 50 multiple-choice questions and short essay questions. However, only the multiple-choice questions were analyzed in this study. Multiple-choice questions can be designed as probes for most of the levels of Bloom's hierarchy (taxonomy) of levels of thinking (Lord and Baviskar 2007), with good examples of comprehension, application, and analysis-type questions appearing on the AP Biology exam, the GRE, and the MCAT (Zheng et al. 2008). I designed multiple-choice questions that required simple recall of facts or comprehension of a concept, as well as application of concepts to a new situation or analysis and interpretation of data or results (see examples in Table 1). Each exam had approximately the same proportion of question types: 50%–60% knowledge, 20%–25% comprehension, 10%–15% application, 5%–15% analysis, typical of many college exams (as reported by Lord and Baviskar 2007). The same format for testing was used in each of the introductory courses I taught from 1994 to 2006, and the exams in different years covered almost exactly the same content with the same proportions of question types, despite the difference in presentation style of the lecture.

Measures and data analysis

Only data from 1998 and 2006 were analyzed in this study, because these two classes were approximately equal in size (46–50 students), and taught at the same time of day, in the same room, and using the same instructional technology. While the two years differed in instructional method, both followed several years of experience with the pedagogy, ruling out instructor experience as a variable in the evaluation of the instructional method on student performance. With such a large time span between the two populations compared in this study, there was the potential for significant differences in student performance based simply on differences in the academic quality of students in the two classes. Some studies have demonstrated that student success in a course is highly correlated with entrance exam scores (Freeman et al. 2007). Correlation analysis indicated that ACT composite score was the best predictor of student success (percentage of total exam points) in both introductory classes ($r = 0.70–0.74$, $p < 0.0001$; Table 2). Using composite ACT scores for each student as the covariate in an analysis of covariance of percentage of total exam points by year (1998 versus 2006), there was no significant difference between the two groups of students ($F = 0.15$, $p = 0.699$). In addition, there was no significant difference between the means of composite ACT scores in the two classes (24.8 $\pm$ 0.56 S.E. for 1998 and 25.8 $\pm$ 0.49 S.E. for 2006; students' t test $p = 0.24$). Therefore, variation in academic ability between the two groups of students did not explain their differences in academic achievement in the classes.

Several different measures of student

TABLE 1

Sample questions from the first exam on evolutionary history of life and biodiversity and last exam on reproductive biology, illustrating knowledge, comprehension, application, and analysis types of responses.

First exam Evolutionary history of life and biodiversity	Knowledge	Gymnosperms differ from ferns in that gymnosperms have • Flowers • Vascular tissue • Seeds • Spores
	Comprehension	Which of the following was NOT a challenge for the first land plants? • Dehydration • Animal predators • Sperm transfer • Support against gravity
	Application	From the Jurassic period to the beginning of the Cenozoic Era, marked differences in the position of the continents occurred through continental drift. Which of the following events might be explained by the changing continental position alone? • Extinction of dinosaurs • Isolation of monotremes in Australia • Radiation of fishes in freshwater • Coevolution of plants and insect pollinators
	Analysis	If forks, knives, and spoons represent predator phenotypes in a landscape where lentils (small), peas (medium), and lima beans (large) represent phenotypes of their prey, what has happened to the predator and prey populations over time, according to the graph below? • Knives evolved into spoons • Lima beans evolved into lentils • The result of natural selection was smaller prey • No evolution occurred in the predators

performance were then analyzed to determine whether instructional method had an impact on student achievement and higher-order thinking skills. All distributions of data for all analyses met the Anderson-Darling test for normality (Minitab 15). As one method of assessing the impact of instructional method on achievement, I compared the point difference between the last and first exams for students in the 1998 and 2006 classes, with the assumption that introductory students would show improvement in test performance as a result of their experience in the classroom and with successive exams. To remove the effect of good student–poor student differences in exam performance throughout the semester, I performed analysis of covariance (using General Linear Models in Minitab 15) on the point difference between last and first exam scores, with composite ACT score as the covariable and year (instructional method) as the main factor.

To determine the effect of instructional method on the development of higher-order thinking skills in these introductory students, I compared how well students answered different types of questions on exams (knowledge, comprehension, application, and analysis). The data analyzed were the percentage of students answering each multiple-choice question *incorrectly* on the first and last exams for both years of the study (the data obtained directly from the Scantron analysis of exam responses). I assumed that if students did improve their higher-order thinking skills throughout the semester, the contrast between first and last exams would be most obvious. I further assumed that lower percentages of incorrectly answered analysis and application questions on last exams compared to first exams would indicate improved skill in higher-order thinking.

I categorized the questions on these exams, assigning them to just two categories for simplification of analysis and to provide increased sample size for statistical comparison: knowledge-

Fourth exam Reproduction of organisms	Knowledge	A baboon troop has one dominant male that does most of the breeding with several females. The best term for this type of reproduction is • promiscuity. • polygamy. • serial monogamy. • polyandry.
	Comprehension	Which of the following is a correct sequence that takes place in flowering plant reproduction? • Meiosis-fertilization-ovulation-germination • Fertilization-meiosis-nuclear fusion-embryo formation • Meiosis-pollination-fertilization-embryo formation • Pollen tube growth-pollination-germination-fertilization
	Application	Why are endocrine-disrupting chemicals more damaging to development of male fetuses than to female fetuses? • Male fetuses are more sensitive to steroids during development. • Endocrine-disrupting chemicals are always estrogen mimics. • Female fetuses don't produce hormones during development, so endocrine disruptors can't harm them. • The male is the derived sex whose development is sensitive to the ratio of androgen to estrogen.
	Analysis	Topminnows living in desert pools in Arizona reproduce both sexually and asexually. A scientist observed that 40% of the asexually reproducing fish were infected with a parasite that caused a disease called blackspot, while few of the sexually reproducing fish exhibited the disease. Following a drought, some of the pools dried up, but were later recolonized by fish from other pools. However, when the scientist checked those recolonized pools, he found that now the sexually reproducing fish had the blackspot disease, while the asexual clones were disease-free. Which of the following might be an explanation for what he observed? • The parasite that causes blackspot attacks only female fish and they passed it on to their offspring. • When fish change their breeding strategy from asexual to sexual they carried the parasite with them. • The pool was recolonized by few sexually breeding fish that were susceptible to the disease. • Asexually reproducing fish populations grow faster than sexually reproducing ones.

comprehension type or application-analysis type. There is justification for making this categorization, as Lord and Baviskar (2007) argue that true higher-order thinking begins with application-and-analysis-type questions. Multiple analysis of variance (with General Linear Models in Minitab 15) was then used to determine the effect of the instructional method on student performance on the two types of multiple-choice questions. This allowed me to test differences in performance based on three factors: question type (knowledge-comprehension versus application-analysis), exam time during the semester (first versus last), and year (lecture versus case-based instruction).

The third measure of student performance I used to compare the two classes was the distribution of total exam points earned by students. In particular, I compared the numbers of students earning greater than 90% of total exam points, 80%-90%, 70%-80%, etc. of the two classes using a Chi-square analysis.

Results

One of the main goals of the study was to determine whether instructional method made a difference in student learning and achievement on exams. Fifty percent of students in the lecture-based (1998) class maintained or improved their scores by an average of 9 points from the first to the last exam, but the other 50% of the class decreased in exam performance, an average of almost 13 points, from first to last exam. In the lecture-based course, there was a significant and positive relationship between the degree of improvement in exam score from beginning to end of the semester and the percentage of total exam points earned over the semester (Figure 1; regression analysis $F = 14.97$, $df = 49$, $p = 0.0003$). In contrast, 80% of students in the case-based class (2006) maintained or improved their score on the last exam compared to the first by a mean of 9 points. Only 20% of students in that class had last exam scores lower than their first exam, and the average decrease was less than three points. In addition, there was no statistically significant relationship between the point difference on last to first exams and the total exam scores for those students ($F = 2.27$, $df = 45$, $p = 0.14$).

Analysis of covariance revealed that instructional method (year) did have a significant effect on the point difference between last and first exams when composite ACT score was used as the covariate in the analysis to remove the effect of students of differing ability ($F = 13.76$; $p < 0.0001$). This suggests that the case-based instruction had a positive and beneficial effect on academic performance, as determined by a differential improvement in exam scores throughout the semester.

The second measure of student performance, the mean percentage incorrect of knowledge-comprehension versus application-analysis type questions, varied significantly between exams and between years. A multiple analysis of variance showed that there were significant effects of year of the study, i.e., instructional method ($p < 0.0001$), exam time during the semester, i.e, first vs. last exam ($p = 0.001$), and category type of question asked ($p < 0.003$) on the percentage of questions answered incorrectly (Table 3). There were no significant interactions of the terms in the analysis.

On the first exam of the semester, students in the lecture-based course (1998) answered knowledge-comprehension type questions as poorly as they did application-analysis-type questions (mean percentage incorrect = 40–41, Table 4), whereas students in the case-based course (2006) answered knowledge-comprehension-type questions reasonably well (mean percentage incorrect = 22) but answered application-analysis questions more poorly (mean percentage incorrect = 41). Thus, at the beginning of the semester the two classes were similar in their lack of ability to process information at the application-analysis level, but differed in their mastery of the content knowledge on the first exam.

On the last exam of the semester, students in the lecture-based course exhibited dramatic and significant improvement in answering knowledge-comprehension type questions (mean percentage incorrect = 29, Table 4), but no significant change in their ability to answer application-analysis-type questions (mean percentage incorrect = 35). In contrast, students in the case-based course exhibited dramatic and significant improvement in answering application-analysis-type questions (mean percentage incorrect = 26) on

TABLE 2

Correlation matrix of ACT test scores and percent of total exam points in introductory biology for 1998 (lecture-based) and 2006 (case-based) classes. All of the correlation coefficients showed significance ($p < 0.01$); coefficients closer to 1.0 represent the strongest correlations. The strongest correlate, and therefore best predictor of the percent of total exam points achieved by students, was the composite ACT score in both classes.

2006	***Exam %***	***ACT-math***	***ACT-English***	***ACT-sci reason***	***act-composite***
Exam %	1				
ACT-math	0.5501	1			
ACT-English	0.5547	0.6476	1		
ACT-sci reason	0.5431	0.6752	0.5431	1	
ACT-composite	0.7366	0.7801	0.8644	0.8192	1
1998	***Exam %***	***ACT-math***	***ACT-English***	***ACT-sci reason***	***ACT-composite***
Exam %	1				
ACT-math	0.6489	1			
ACT-English	0.4511	0.7139	1		
ACT-sci reason	0.5444	0.7236	0.6849	1	
ACT-composite	0.6968	0.6489	0.8110	0.8273	1

FIGURE 1

Improvement in exam scores (last exam points – first exam points) in introductory biology students as a function of their cumulative performance in the course (percentage of total exam points). Improvement in exam performance (+ difference) was significantly related to overall performance (regression *F* = 14.97, *df* = 49, *p* < 0.0001) in the lecture-based course (1998), but not in the case-based course (2006).

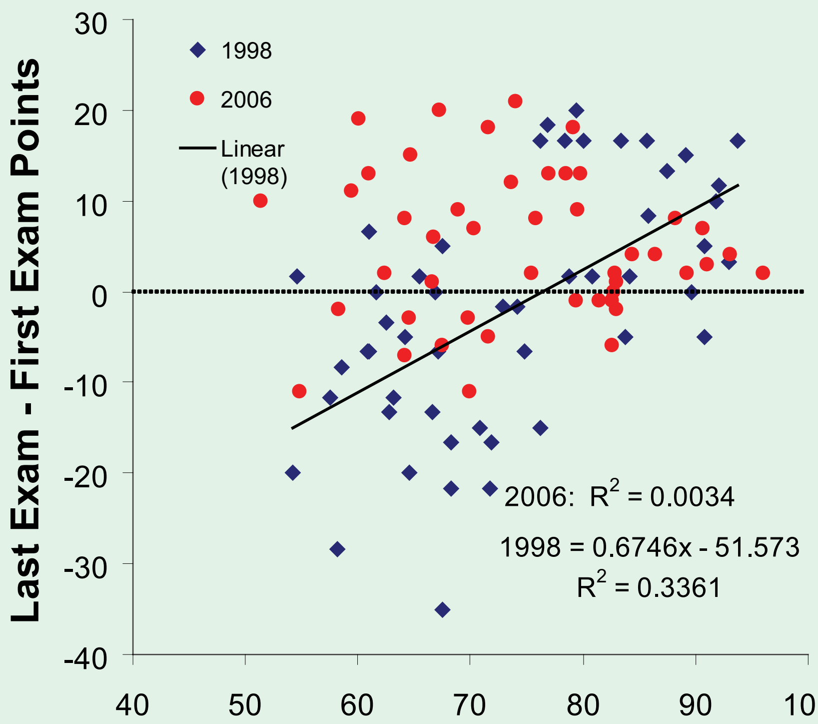

TABLE 3

Results of the General Linear Models analysis of percentage *incorrect* responses on exam questions in introductory biology as a function of three factors: year (lecture-based teaching in 1998 versus case-based teaching in 2006), the exam time during the semester (first exam versus last exam), and the category type of question (knowledge-comprehension versus application-analysis).

Source	*F*	*p*
Year	*11.84*	*0.001*
Exam time	12.22	0.001
Type of question	8.92	0.003
Year * Type	2.8	0.097
Exam * Type	1.25	0.266
Year * Exam	0.67	0.414

the last exam. Thus, students in the case-based course exhibited mastery of content to the same degree that students in the lecture-based course did, but also exhibited improved ability to process information using higher-order thinking on the last exam of the semester.

The last measure of student performance examined in this study was the distribution of the percentage of total exam points earned by students. In 2006, 40% of students in the case-based course earned at least 80% of the total exam points, compared to only 30% of students with that many points in lecture-based course in 1998 (Figure 2). There were still 29% of students in the case-based course who earned less than 70% of the exam points, but that was markedly less than the 46% of students in the lecture-based course who achieved at that level. Despite this perceived improvement in performance of the case-based class, a Chi-square test of the distributions of total exam percentages in each decadal bracket (50, 60, 70, etc.) revealed no significant difference in the distribution of scores by decadal bracket between the two classes (Chi-square = 3.533; *df* = 4; *p* = 0.473).

Discussion

Although a number of studies report increased learning gains by students involved in various sorts of inductive (involving active learning) exercises in the classroom (Dinan 2002; Klionsky 2001–2002; O'Sullivan and Copper 2003; Prince 2004; Lawrenz, Huffmann, and Appeldoorn 2005; Allen and Tanner 2005; Knight and Wood 2005; K.A. Smith et al. 2005; Camill 2006; Freeman et al. 2007), they use a wide variety of teaching and learning strategies, not necessarily just case studies, and few of them provide quantitative proof of the learning gain. In a review of inductive teaching and learning methods, Prince and Felder (2006) cite only three studies that report quantitative evidence of gain in students' higher-order critical-thinking skills with the use of case studies in the science classroom, all of which were papers presented at meetings. The goal of this study was to specifically examine the impact of a transition from lecture-based to case-based teaching on student performance and development of the cognitive skills that characterize higher-order thinking, specifically ability to apply their knowledge and analyze novel situations.

In this study there was a significant and marked impact of instructional method on student performance, based on the amount of improvement in exam scores from first to last exam. Not only did more students in the case-based class improve from first to last exam (80% compared to 50% in the lecture-based class), but those whose scores fell showed only a minor decline (an average of less than 3 points), compared to the average major decline of 13 points in the lecture-based class. Further, students in the case-based course showed more improvement over the course of the semester, as well as slightly higher overall success at answering application-analysis-type questions correctly than students in the lecture-based course (Table 4). These data support the contention by some (Halpern 1998; Prince 2004; Lauer 2005) that the development of higher-order thinking skills can be facilitated when students have practiced by grappling with complex, open-ended questions, applying their content knowledge to the novel situations posed by case studies, and developing their own strategies for problem solving.

It is interesting to note that in this study, students in the lecture-based course did improve markedly in their ability to answer knowledge-comprehension-type questions over the semester (Table 4), compared to the more constant, but high level of performance of students in the case-based course on these kinds of questions. Case studies provide context and a strong stimulus and purpose for learning content. It could be argued that by organizing content learning around the theme or problem of the case, students might be able to master and recall vocabulary and concepts more easily, even as be-

FIGURE 2

Distribution of total exam points for students in a lecture-based course (1998) compared to a case-based course (2006). Percentage of students achieving a particular percentage of the maximum possible exam points is shown for five ranges. Differences in the distributions were not significant by Chi-square analysis (*p* = 0.473).

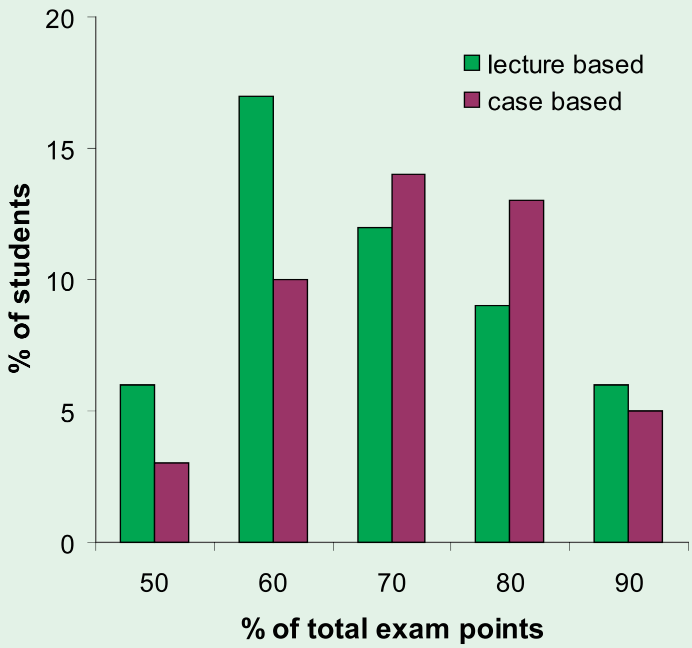

TABLE 4

Individual means (standard error) for percentage *incorrect* responses to knowledge-comprehension and application-analysis-type questions on first and last exams in introductory biology in lecture-based (1998) versus case-based (2006) classes.

Year	First exam		Last exam	
	Knowledge-comprehension	Application-analysis	Knowledge-comprehension	Application-analysis
1998	40 (3.0)	40.8 (3.68)	28.9 (3.49)	34.5 (5.43)
2006	22.1 (2.48)	40.8 (3.43)	21.8 (2.29)	25.8 (2.82)

ginning students early in the semester. Yadav et al. (2007) report that faculty perceive that their students do learn content more deeply and exhibit better exam performance when they have studied new material in the context of a case study.

In this study, there was a trend toward an increased B-level performance and decreased D- and F-level performance, with the transition to case-based presentation style. Because the class size and student ability, instructor and instructor experience, text, exam format and question types, grade scale, and content organization of the courses were similar for both groups of students, the variable of instructional method may explain the measured differences in student performance. Knight and Wood (2005) reported a similar trend of improved student performance (percentage of total course points) when they transitioned from entirely lecture to an interactive environment in the classroom. In addition, they documented higher learning gains by students in the interactive environment, based on the difference between responses to multiple-choice questions given at the beginning of the semester and later embedded in the final exam. One would hope this early success in an introductory-level course would encourage students to pursue more courses in the discipline where their development of critical-thinking skills (application and analysis) would serve them well.

References

Allen, D., and K. Tanner. 2005. Infusing active learning into the large-enrollment biology class: Seven strategies, from simple to complex. *Cell Biology Education* 4 (4): 262–68.

Camill, P. 2006. Case studies add value to a diverse teaching portfolio in science courses. *Journal of College Science Teaching* 36 (2): 31–37.

Campbell, N.A., and J.B. Reece. 1996, 2004. *Biology*. 4th and 7th eds. Upper Saddle River, NJ: Addison Wesley and Pearson.

Dinan, F.J. 2002. Chemistry by the case. *Journal of College Science Teaching* 32 (1): 36–41.

Freeman, S., E. O'Connor, J.W. Parks, M. Cunningham, D. Hurley, D. Haak, C. Dirks, and M.P. Wenderoth. 2007. Prescribed active learning increases performance in introductory biology. *CBE—Life Sciences Education* 6 (2): 132–39.

Halpern, D.F. 1998. Teaching critical thinking for transfer across domains. *American Psychologist* 53 (4): 449–55.

Handlesman, J., D. Ebert-May, R. Beichner, P. Bruns, A. Chang, R. DeHaan, J. Gentile, S. Lauffer, J. Stewart, S.M. Tilghman, and W.B. Wood. 2004. Scientific teaching. *Science* 304 (5670): 521–22.

Klionsky, D.J. 2001–2002. Constructing knowledge in the lecture hall. *Journal of College Science Teaching* 31 (4): 246–51.

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

Lauer, T. 2005. Teaching critical-thinking skills using course content material: A reversal of roles. *Journal of College Science Teaching* 34 (6): 34–37.

Lawrenz, F., D. Huffman, and K. Appeldoorn. 2005. Enhancing the instructional environment. *Journal of College Science Teaching* 34 (7): 40–44.

Lord, T., and S. Baviskar. 2007. Moving students from information recitation to information understanding: Exploiting Bloom's taxonomy in creating science questions. *Journal of College Science Teaching* 36 (5): 40–44.

Lundeberg, M.A., and A. Yadav. 2006a. Assessment of case study teaching: Where do we go from here? Part I. *Journal of College Science Teaching* 35 (5): 10–13.

Lundeberg, M.A., and A. Yadav. 2006b. Assessment of case study teaching: Where do we go from here? Part II. *Journal of College Science Teaching* 35 (6): 8–13.

O'Sullivan, D.W., and C.L. Copper. 2003. Evaluating active learning. *Journal of College Science Teaching* 32 (7): 448–52.

Prince, M. 2004. Does active learning work: A review of the research. *Journal of Engineering Education* 93 (3): 223–31.

Prince, M., and R.M. Felder. 2006. Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education* 95 (2): 123–38.

Schiller, N.A., and C.F. Herreid. 1999. The Galapagos. The National Center for Case Study Teaching in Science Case Collection. *http://ublib.buffalo.edu/libraries/projects/cases/galapagos1.html* .

Smith, A.C., R. Stewart, P. Shields, J. Hayes-Klosteridis, P. Robinson, and R. Yuan. 2005. Introductory biology courses: A framework to support active learning in large enrollment introductory science courses. *Cell Biology Education* 4 (2): 143–56.

Smith, K.A., S.D. Sheppard, D.W. Johnson, and R.T. Johnson. 2005. Pedagogies of engagement: Classroom-based practices. *Journal of Engineering Education* 94 (1): 87–101.

Yadav, A., M. Lundeberg, M. DeSchryver, K. Dirkin, N.A. Schiller, K. Maier, and C.F. Herreid. 2007. Teaching science with case studies: A national survey of faculty perceptions of the benefits and challenges of using cases. *Journal of College Science Teaching* 37 (1): 34–38.

Zheng, A.Y., J.K. Lawhorne, T. Lumley, and S. Freeman. 2008. Application of Bloom's taxonomy debunks the "MCAT Myth." *Science* 319 (5862): 414–15.

Susan Chaplin (sbchaplin@stthomas.edu) *is a professor of biology at the University of St. Thomas in St. Paul, Minnesota.*