

Article

Cloning the Professor, an Alternative to Ineffective Teaching in a Large Course

Jennifer Nelson,* Diane F. Robison,[†] John D. Bell,* and William S. Bradshaw[‡]

Departments of *Physiology and Developmental Biology; [†]Instructional Psychology and Technology; and [‡]Microbiology and Molecular Biology, Brigham Young University, Provo, UT 84602

Submitted January 26, 2009; Revised June 4, 2009; Accepted June 17, 2009

Monitoring Editor: Paul Williams

Pedagogical strategies have been experimentally applied in large-enrollment biology courses in an attempt to amplify what teachers do best in effecting deep learning, thus more closely approximating a one-on-one interaction with students. Carefully orchestrated in-class formative assessments were conducted to provide frequent, high-quality feedback that allows students to accurately diagnose the current state of their understanding of fundamental biological concepts and make specific plans to remedy any deficiencies. Teachers can also assume responsibility to guide out-of-class study among classmates by promoting Elaborative Questioning, an inquiry exchange that permits misconceptions to be identified and corrected and that promotes long-lasting metacognitive and analytical thinking skills. Data are presented that demonstrate the positive impact of these innovations on student performance and affect.

INTRODUCTION

What is the cost of conducting a single session of your course? Making that estimate is instructive, so let's all pick up a pencil and perform the calculation using Table 1 as a helpful template. To bring one of us (Bradshaw) and the 175 students enrolled in Biology 360 (Cell Biology, Fall 2007) together in the same room for a single class period cost more than \$5000. What dollar value did you arrive at for your course? Like us, you're probably surprised at the size of the figure. And what conclusion do we draw, then, from this demonstration that the logistic cost of meeting our classes each time is very expensive?

We suggest that the message is, "Don't carry out a learning exercise in the classroom setting that can be accomplished more effectively in another venue." Alternatively, "In trying to promote meaningful learning, a teacher should use each available resource for that purpose for which it is best suited." Serious deliberation of these assertions will be helpful because it leads us to identify what we do best as teachers, and how to use our particular strengths to advantage. Here, we report one means of accomplishing this goal: finding various ways to "clone" those teacher functions that

will provide the greatest benefit to the greatest number of students.

Unfortunately, the vision of a single, eager learner sitting at the feet of Socrates can only be an unrealizable ideal. Because of the immense cost required, we don't have the luxury of providing what in the educational realm would be analogous to a personal trainer helping a client to improve physical fitness. The dilemma leaves us with two options: maintain the status quo (doing the best we can with large enrollments); or, as we propose, identify creative ways to amplify or clone what a teacher does in such a way as to approximate the one-on-one scenario.

The first cloning strategy might be to modify the course design so that, beyond traditional office hours, opportunity for individual attention to a small number of students is a regular part of the pedagogy. Alternatively, the number of such opportunities can be expanded by recruiting helpers who are capable of providing the equivalent high-quality mentoring service. The second technique is to widen the effective influence of the teacher when he or she is physically present with the students. In this mode, the teacher models the desired intellectual activity with one or a few students and others benefit by engaging vicariously in the demonstration. The third approach is for the teacher to program learning activities outside of class, when he or she is not present, in such a way that students interacting with one another mimic (perhaps imperfectly at first) the peda-

DOI: 10.1187/cbe.09-01-0006

Address correspondence to: William S. Bradshaw (william_bradshaw@byu.edu).

Table 1. Economics of conducting one class period in a course

	Estimated costs	My course	Biology 360, \$
1	Value of 1 h of teacher's time in class		40.00
2	Teacher before class preparation time (2 h)		80.00
3	Value of student's time (no. of students \times \$8, estimated hourly wage)		1400.00
4	Students before class preparation (1 h)		1400.00
5	Teacher expense: commuting to the room		7.00
6	Student expense: commuting to the room (no. of students \times \$5)		875.00
7	Tuition: divide the tuition cost/term by 200 and then multiply by the number of students		1584.00
8	Teaching assistants (number \times hourly wage)		24.00
9	Additional costs (e.g., supplies, equipment rental)		20.00
10	Overhead costs		50.00
11	Total (lines 1–10)		5480.00

gological expertise of the teacher. We report here the implementation of four pedagogical techniques derived from these theoretical possibilities (formative assessment, elaborative questioning, faculty mentoring sessions, and alumni consultation), and we provide both performance and affective data demonstrating their effectiveness.

Adopting a "cloning the professor" strategy is a response to our recognition of two common student weaknesses: a shortsighted perspective on the most meaningful benefits of education and a superficial approach to the methodology of learning.

Too often students focus solely on passing exams and receiving high grades as a narrow route to becoming credentialed. In Table 2 we propose what we hope is a more essential set of student necessities along side a short, generic

catalog of those personal teacher resources that might be marshaled to help meet those needs. Although these kinds of educational objectives are more difficult to achieve, they are also more likely to help people attain the long-term personal and professional goals implicit in the desire for a high GPA. In fact, they parallel quite closely educational goals identified in large-scale surveys of parents (Goodlad, 1984), the preferences of college seniors (Light, 2001) and college faculty (Bok, 2006), and are among those attributes deemed most useful for success in a future career (Overtoom, 2000, see especially "Adaptability Skills"; Bok, 2006).

Generally, most students are not very adept at meeting the needs listed in Table 2 by themselves; making progress toward achieving those ends is likely to be slow without the intervention of a good teacher. The reason for this is that in addition to being novices with respect to the specific knowledge and skills pertaining to a particular academic discipline, beginning college students also tend to be novices with respect to the fundamentals of knowing how to learn. They have often been able to get by in earlier phases of their education, in part because the superficial study strategies they routinely used were sufficient to satisfy course requirements that lacked adequate rigor. "Students succeed," agree McClymer and Knowles (1992), "because despite the fact that the techniques and strategies they employ have only incidental connection to the disciplines involved, they are exceedingly well adapted to the sorts of tasks we ask them to perform in our courses." The result, they continue, is "ersatz [spurious] learning." Light (2001) echoes this view, stating that many students will struggle who "continue to organize their work in college the same way they did in high school." Especially at risk, he says, are "those who were academic stars in high school, but at schools that made only modest demands upon them." Some common symptoms of these weak learning strategies include poor reading skills, acquiring and retaining misconceptions about basic concepts, studying in isolation rather than actively engaging others, and depending on short-term memorization without assessing deep understanding. We return to these problems in the *Discussion*, and in a sometimes anecdotal style, we illustrate how they might be ameliorated through the cloning techniques described and assessed below.

MATERIALS AND METHODS

Description of Courses

Biology 120 is the first course (3 credit hours) in a core curriculum that serves students enrolled in a variety of biology-related majors. Its design differs from that of the traditional broad survey offering: a relatively small number of foundational concepts are introduced (membrane and organelle function, the central dogma of molecular biology, genetics, energy transformations, and evolution), with emphasis on a thorough understanding of basics at the expense of detail. In addition, students are expected to master some aspects of the conduct of experimental biology and of data analysis, e.g., controls, dependent and independent variables, distributions, *p*-values, and correlation. The text used was Freeman, *Biological Science*, 2nd edition. In winter semester 2006 when most of the data reported below were gathered, there were 263 students enrolled, 79 females and 184 males. The four class levels were represented as follows: 102 freshman, 94 sophomores, 48 juniors, and 16 seniors.

Biology 360, Cellular Biology, is an upper-level 3-credit-hour offering in the same core curriculum. The stated aim of the course is

Table 2. Comparison of what students require and what teachers can provide

Student needs	Teacher assets
Correct misconceptions	Subject matter expertise
Develop intellectual passion for important ideas	Personality (potential to provoke, motivate, and inspire)
Think like practitioners in the field	Research experience (academic maturity)
Generate a self-confident disposition to address and solve problems	Successful pedagogy (rigorous inquiry exercises)
Acquire inquiry and metacognitive skills for life-long learning	Diagnostic skill (ability to discern learning weaknesses in novices)

for students to develop facility in the interpretation of experimental results, specifically to be able to construct coherent sentences that correctly set forth conclusions supported by the data (Kitchen *et al.*, 2003). The course has a reputation for academic rigor; short-term memorization of factual information will not suffice. The text used was Alberts, *Molecular Biology of the Cell*, 4th edition. Across the nine semesters during which data were collected (winter 2002–fall 2007), the average enrollment in Biology 360 was 156 (range, 94–206). Typically, the course enrolls 20% juniors and 80% seniors. Approximately 35% of the class members are women. None of the students represented in the Biology 120 data set also was represented in the Biology 360 set.

Pedagogical Strategies

Formative Assessment. Formative assessment meets the need for a student to expose his or her understanding to scrutiny with the intent to improve performance. It is a cloning vehicle because it disseminates high-quality professorial feedback. In emphasizing the importance of this kind of diagnostic evaluation, Fink suggests that feedback must be of the “FIDeLity” variety, that is, frequent, immediate, discriminating (based on criteria and standards), and done lovingly (Fink, 2003, p. 95).

We have introduced the use of formative assessments in two different ways: the fully formative assessment model and the hybrid assessment model (Figure 1). In the fully formative assessment model, all traditional midterm exams are eliminated (Kitchen *et al.*, 2006). Instead, weekly formative assessments are administered in class under test-like conditions. Students are given 20–25 min to answer two to three questions designed to help them assess their understanding of key course concepts as well as their proficiency in data analysis. Next, students are asked to share and critique their answers with each other. Then the answers are revealed and the professor models how to think about the questions so as to arrive at the correct conclusions. Students can ask questions and seek further clarification from the professor and teaching assistants, who roam the aisles interjecting themselves into small group discussions. The room is very noisy. Students may receive points for their participation but not their performance. This provides a low-risk environment in which every student can make a personalized evaluation of his or her understanding and move in the direction of mastery. At the conclusion of the course, the final exam is administered. It is the only summative assessment.

The hybrid assessment model is a combination of the formative assessment model and the traditional summative model. Formative assessments are interspersed between traditional summative exams. Quantitative evidence suggests that both models are very effective in promoting mastery of basic concepts and improved analytical thinking (Kitchen *et al.*, 2003, 2006; this study).

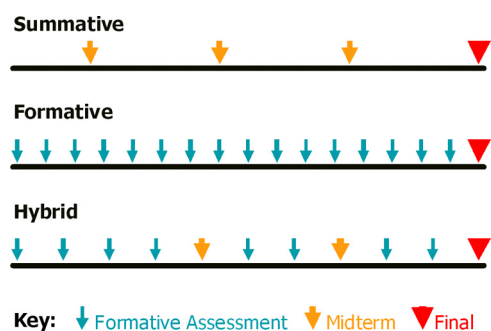


Figure 1. Course design options for introducing formative assessment. Key: formative assessment, blue-green arrows; midterm exam, orange arrowhead; final exam, red inverted triangles.

Elaborative Questioning. Learning is improved when students engage each other in conversation. Tanner (2009) describes the benefits of “Student Talk” in the classroom and lists ways in which this activity can be implemented. Talking also improves the effectiveness of out-of-class study. In preparation for exams, most students study alone, reread the text, and review their notes. This often promotes short-term rote retention of the subject matter. To promote genuine ownership and deeper understanding, we have encouraged students to modify how they study by requiring them to participate in at least 1 h of Elaborative Questioning (EQ) a week. EQ is a study strategy that is a modification of Mark McDaniel’s concept of elaborative interrogation (EI). McDaniel found that students who responded in written form to “why” questions after reading textbook material performed better on an assessment than their peers who just read the text (McDaniel and Donnelly, 1996). The underlying principle of EQ is the same as EI, but the exercise is carried out in significantly different circumstances. Students get together after class in small groups and generate their own “how” and “why” questions, and then critique the answers that are given. They take turns as both the questioner and responder. To introduce the activity and help members of the class understand what constitutes a productive EQ session, they are shown a short video in which the process is modeled by former students. In addition, faculty continually model EQ and promote its use in their in-class presentations. EQ becomes a cloning device, then, because students learn to mimic a teacher-student dialogue with the attendant benefits that would accrue if the teacher were actually present.

Faculty Mentoring Sessions. We have introduced opportunities into the academic week of Biology 120 and Biology 360 for students to interact directly with the faculty in addition to the regularly scheduled class sessions. These mentoring sessions, scheduled up to five times per week, are devoted primarily to practicing the types of analytical assessment problems that students find most challenging, for example, those requiring interpretation of experimental data. In a typical session, students interact with one another in groups of two to four as they work out solutions to these homework problems. This occurs under the direction of the faculty instructor who formulates leading questions, participates in the small group discussions, and provides feedback, both individually and to the group as a whole. Students normally submit their answers to the homework problems, which are graded, but attendance at the mentoring sessions is not recorded or awarded with points. We estimate that 70–90% of students enrolled in these courses attend at least one of these 1-h sessions each week.

Alumni Consultation. We have experimented with another resource that has significant potential for cloning teacher-like functions—course alumni (Biology 360) who have excelled in prior semesters. We have enlisted the services of these undergraduates who volunteer, without pay, to spend 1 h each week helping class members in the current semester. Participation was voluntary; points were not awarded. However, specific persons who were thought to be in need of assistance were issued individual invitations. Table 3 shows how these sessions were used. Approximately 65% of the students registered in the course attended at least once, including 14 of those who ranked in the top 20 overall at the end of the semester. Several who struggled with the data analysis task, and ranked near the bottom at the end of the semester attended these sessions multiple times. The information, advice, and encouragement these former students provided was superb, and of similar quality, we believe, to what we as teachers would give if circumstances permitted.

Data Collection

Performance in Biology 120 was assessed using 22 multiple-choice questions administered optionally (163 students responded) during the first week in the semester and again as part of the final exam. Course evaluation data were collected anonymously (138 respon-

Table 3. Participation in "alumni consultation" help sessions

Visits/student	No. of students enrolled in course (%)
0	44 (36.4)
1	30 (24.8)
2	18 (14.8)
3	13 (10.7)
4	11 (9.1)
5	1 (0.8)
6	1
7	1
9	1
12	1

Biology 360, Cell Biology, fall 2007.

dents) at the end of the semester. In addition, final exam data were used from multiple semesters between 2004 and 2008.

In Biology 360 a standardized rubric was created for use by raters grading exam problems that assessed analytical reasoning. The uniform task was "State in one sentence each of the conclusions validated by the data presented herein." The performance results were tabulated over a sequence of nine semesters, five before (751 total students) and four after (662 total students) the introduction of formative assessment class periods. Data on the affective responses of students to the design and management of the course were obtained at various intervals during the semester through questionnaire items included as part of each formative assessment.

This project was reviewed by Brigham Young University's Institutional Review Board (IRB) and given exempt status. Informed consent was not required by the IRB as long as data were reported only in aggregated form with no possible link to individual students.

RESULTS

Impact of Formative Assessment Strategies on Academic Performance

In one semester of Biology 120 that used the fully formative model, student performance improved significantly on a set of 22 multiple-choice test items administered in a pretest-posttest format (Table 4). Improvement on 21 of the items ranged from 13 to 67%, with an average gain across all 22 items of 33%. The questions covered a broad range of fundamental curriculum concepts and required cognitive skills ranging from basic comprehension to application and analysis. Two representative problems from this set are shown in Figure 2. The data in Table 5 suggest that the introduction of EQ and the fully formative pedagogy during fall semester 2005 were the most significant factors promoting the observed performance gains. Final exam scores increased abruptly in that year and thereafter remained higher in comparison with the previous three semesters before these interventions were incorporated.

Several lines of evidence also indicate that students benefit significantly from the hybrid formative model used in Biology 360, Cell Biology. A pretest-posttest method of assessment was also effective in demonstrating the acquisition of improved analytical thinking skills (drawing valid conclusions from experimental data) over the course of a semester. The data in Table 6 show that there was an average gain

of 17% on problems covering nonbiological topics (that did not require special experimental expertise or technical vocabulary) and an average gain of 46% on problems whose data were closely aligned with the subject matter of cell biology.

The results of a direct test of the efficacy of the hybrid assessment model in promoting improved performance compared with the traditional summative model are shown in Figure 3. Scores on examination problems administered in five prior semesters by using the traditional summative model are compared with scores on the same problems from a set of four semesters after the hybrid assessment model was introduced. The novel use of formative assessments was the only significant variable between these two sets of trials; the subject matter addressed, classroom presentations by faculty, classroom inquiry exercises, and required homework problems were identical. The problem whose scores are reported in Figure 3A required a calculation of membrane potential. It served as an internal control because the formulaic nature of the task was not expected to improve significantly through formative practice: there was no significant difference in the scores generated through the two formats. In contrast, the hybrid formative assessment format generated a performance gain of 10.5% on a set of data analysis problems in the first midterm exam (Figure 3B), which was maintained (12%) in midterm exam 2 (Figure 3D). Moreover, this improvement (8%) extended to a representative conceptual problem (one assessing general knowledge, including modeling, of regulation of the cell cycle; Figure 3C).

Impact of Elaborative Questioning

Table 7 reports how Biology 360 students' attitudes about EQ as a learning resource were related to overall class standing (from 1, first, to 123, last) as a measure of performance. Although scores on several assignments contributed, class rank depended most heavily on achievement on exams, two midterms, and the final. When queried at the half-way point of the semester, ~84% of respondents viewed EQ as helping them perform better on the first midterm exam. The average score on the final exam for this group was 17% higher than those whose views about EQ were negative, and their ranking in the course was an average 20 persons higher. Those whose sample EQ questions were judged to be of highest quality scored ~28% better on the final and ranked ~32 people higher in final standing, suggesting a positive connection between EQ and academic performance. The same general trends were apparent when the measure was the intent to continue to use EQ in future courses (94% of the responses were affirming). These latter data also suggest the existence of a very small cohort of students who are either independently capable of a strong performance without EQ or do not respond to its benefits (i.e., 5.6% indicated that EQ was not helpful).

Impact of Formative Assessment and Elaborative Questioning on Affect

Overall, the course design and pedagogical methods used in both the introductory and advanced courses described here were received very positively by students. The data in Table 8 indicate that 87% of the students in Biology 360 and 88% of

Table 4. Performance gains in an introductory biology course by using a fully formative assessment design

	Topic	Proportion correct		Gain	Cognitive process
		Pretest	Posttest		
1	Central dogma	0.38	0.51	0.13	Analysis
2	Central dogma	0.58	0.57	-0.01 ^a	Application
3	Genetics	0.34	0.48	0.14	Comprehension
4	Experimental design	0.51	0.66	0.15	Analysis
5	Experimental design	0.26	0.59	0.33	Application
6	Experimental design ^b	0.20	0.56	0.35	Analysis
7	Respiration and photosynthesis	0.34	0.58	0.25	Analysis
8	Mitosis/meiosis	0.33	0.88	0.55	Comprehension
9	Respiration and photosynthesis	0.21	0.88	0.67	Comprehension
10	Mitosis/meiosis	0.42	0.80	0.38	Comprehension
11	Mitosis/meiosis	0.61	0.93	0.32	Comprehension
12	Genetics	0.67	0.94	0.28	Analysis
13	Genetics	0.47	0.72	0.25	Application
14	Central dogma ^b	0.13	0.76	0.63	Application
15	Genetics	0.37	0.78	0.41	Analysis
16	Evolutionary theory	0.64	0.94	0.29	Application
17	Experimental design	0.63	0.87	0.24	Application
18	Cell	0.46	0.88	0.42	Comprehension
19	Experimental design	0.33	0.94	0.62	Application
20	Experimental design	0.33	0.81	0.48	Application
21	Cell	0.82	0.99	0.17	Comprehension
22	Evolutionary theory	0.77	0.92	0.15	Comprehension

Biology 120, winter 2006, $n = 263$. The items are ordered on a scale of relative difficulty (>0.7 , easy; <0.7 , difficult): items 1–7, difficult on both pre- and posttests; items 8–20, difficult on pretest, easy on posttest; items 21–22, easy on both pre- and posttests. Identification of the appropriate cognitive process involved in each problem was based on Bloom's taxonomy (Anderson and Krathwol, 2001).

^a $p < 0.01$ for the gain on all problems, except 2 (two-tailed Student's t test for proportions).

^b The text for items 6 and 14 is shown in Figure 2.

the students in Biology 120 perceived that the benefits they received were worth the rigorous scholastic demands of the experience. Ninety-one percent of cell biology students felt

Item 6 (Data Analysis)

Cells in Dr. Bell's research laboratory were treated with ionomycin, a drug that causes cellular stress. The membranes on these cells were then tested to see if the stress reduced ATP levels. Group A received no ionomycin. Group B received ionomycin. Group C received vanadate, a drug known to lower ATP levels. The data obtained from 4 samples were as follows:

Group A (mM ATP):	1.4, 1.6, 0.9, 2.0
Group B (mM ATP):	1.1, 0.0, 1.3, 1.5
Group C (mM ATP):	0.1, 0.2, 0.1, 0.3

A reasonable conclusion based on these data is that

- | | |
|---------------------------------|---|
| A. vanadate lowers ATP levels. | D. vanadate increases ATP levels. |
| B. stress lowers ATP levels. | E. vanadate does not alter ATP. |
| C. stress increases ATP levels. | F. stress does not alter ATP levels. |

Item 14 (Central Dogma)

A certain protein consisting of a single polypeptide chain has a molecular weight of approximately 31,000. What is the best estimate for the minimum molecular weight of the gene (the double-stranded segment of DNA) which programs the synthesis of this protein? The molecular weight of an average amino acid is 120 g/mol, and the average nucleotide molecular weight is 310 g/mol.

- | | | | |
|------------|------------|-------------------|------------|
| A. 72,000 | D. 188,000 | G. 480,000 | J. 717,000 |
| B. 96,000 | E. 236,000 | H. 558,000 | |
| C. 124,000 | F. 393,000 | I. 603,000 | |

Figure 2. Examples of pre-post assessment items.

that the emphasis on learning strategies and becoming more metacognitive was helpful. When queried more specifically about the value of EQ and formative assessments, the responses were similarly favorable: $>85\%$ in each course perceived benefits from engaging in EQ and 95% of cell biology students felt that formative assessments had helped them to improve their skill in analyzing data. (Students in Biology 120 were not surveyed on this issue.) It is encouraging that students' motivation seemed to stem from a genuine desire to understand the subject (3.4 on a scale from 0 to 5) rather than treating EQ as just another course requirement (2.7 on a scale from 0 to 5). Students also seem to have made an effort to make sure everyone derived benefit from participating actively during EQ sessions (4.4 on a scale from 0 to 5).

Table 5. Longitudinal assessment of final exam performance in an introductory biology course

	Before "cloning" interventions	After "cloning" interventions
Class avg (%)	72.0 ^a \pm 3.6 ($n = 3$)	76.8 ^b \pm 2 ($n = 6$) ^c

^a 2004–2005.

^b 2005–2008.

^c $p = 0.02$, Student's t test.

Table 6. Performance gain in data analysis skill using a hybrid formative assessment course design in Biology 360

Data source	Problem	Semester	n	Pre	Post	Gain (%)
Generic	Chess	W06	193	4.70	5.51	17†
		F06	124	4.30	5.13	19†
		W07	157	5.11	5.96	17†
	Transfer	F07	222	6.48	7.41	14*
Biology	Secretion	W06	196	3.27	5.34	63†
		F06	124	3.66	5.00	27†
	Prolactin	W07	156	3.20	5.10	38†
		F07	222	4.30	6.77	57†

† $p < 0.001$ and * $p < 0.05$, two-tailed unpaired t test.

Raw scores from presemester and postsemester measures of data analysis skill. The task was uniformly to write sentences that correctly stated conclusions validated by the data presented in the problems. Total points possible: 10 (chess, prolactin, and secretion); 15 (sheep and transfer). Instruction in the four semesters used the hybrid formative format in Biology 360. The statements that follow represent optimal responses to the exam items: Chess (Chase and Simon, 1973). Chess masters have a greater capacity to detect patterns than do chess novices; Transfer (Michael *et al.*, 1993). The "anchored instruction" pedagogy was more successful than traditional instruction in helping students achieve transfer on examination tasks in a speech-language pathology course; Secretion (J. D. Bell, fictitious data). The protein is a secretory protein that follows the pathway from rough endoplasmic reticulum to Golgi to extracellular matrix; Prolactin (Guyette *et al.*, 1979). Prolactin promotes transcription of the casein gene and increases the half-life of the casein mRNA; Sheep (Roselli *et al.*, 2004). Male-oriented rams have reduced hypothalamic structures and reduced levels of aromatase compared with female-oriented rams.

Survey results also indicate that EQ quickly became integrated into the study routine of the Biology 360 course and was perceived as a very valuable element. The data in Table 9 suggest that students were engaging in those activities that were more likely to be genuinely helpful in preparing for the first midterm exam (hybrid formative format). Ninety-five percent of students planned to engage in an EQ session; it was deemed to be among the most productive study activities. Encouragingly, only 42% were planning to reread the pages of the text for this purpose.

Students are also highly positive about their experience in faculty mentoring sessions. In the winter semester 2006, 198

students (94% of those enrolled) in Biology 360 responded to the following survey question: "Which of the following represents your opinion of the value of the weekly Mentoring Sessions?" Given four choices, 76.8% selected "Very Helpful" and an additional 18.7% selected "Somewhat Helpful." The average score on a Likert scale assessment of faculty mentoring (0–5, where 0 indicated not beneficial and 5 indicated highly beneficial) for 176 students (98% of those enrolled) in the winter 2007 semester of the same course was 4.3. As indicated in Table 3, 63% of the fall semester 2007 students in Biology 360 availed themselves of consultation opportunities with course alumni, many on multiple occa-

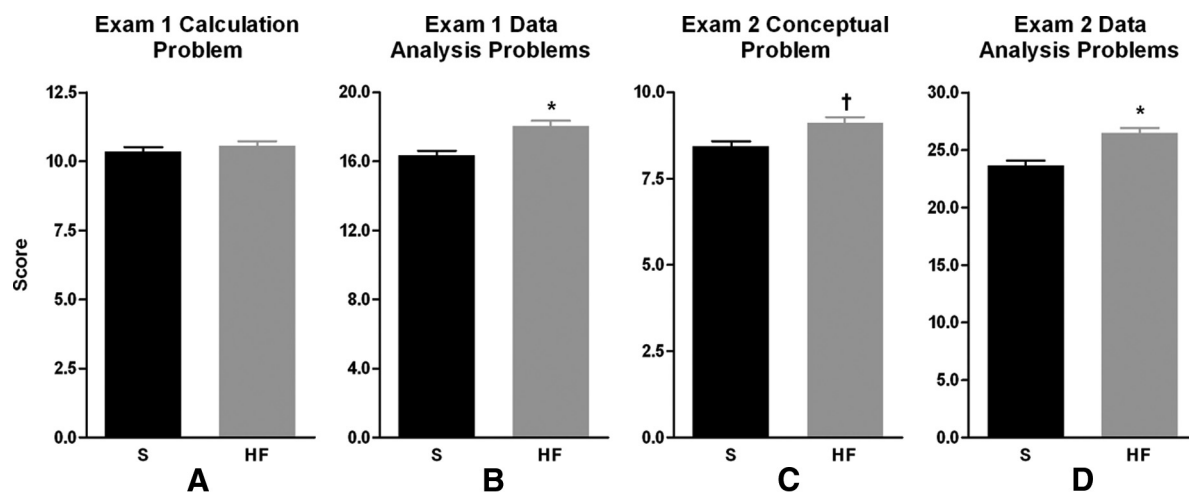


Figure 3. Performance gain on conceptual and data analysis problems in cell biology (Biology 360). S, summative course design (five semesters, $n = 662$). HF, hybrid formative course design (four semesters, $n = 751$). The exam 1 data analysis set consisted of two problems. The exam 2 data analysis consisted of three problems. The calculation and conceptual sets consisted of a single item each. Fifteen points were possible on all problems. For each problem set, data from the HF design were compared with the S design using a two-tailed two-sample t test: * $p < 0.0001$ and † $p < 0.002$. Error bars represent the SEM.

Table 7. Relationship between participation in elaborative questioning and course achievement

Index of EQ validity	Response	%	Avg. final score	Avg. final rank
Perception of EQ efficacy ^a	Yes	83.5	90.0‡	54.4§
"Did your participation in earlier EQ sessions actually help you perform better on the first midterm exam?"	No	16.3	76.6	75.4
Quality of EQ questions: ^b students submitted sample questions that reflected the best that had emerged during a recent EQ session. These were evaluated as to their ability to stimulate deep learning.	None	11.0	85.6	64.8
	Poor	25.7	78.9	72.4
	Average	37.6	83.6	61.7
	Good	25.7	103.7*	34.6†
EQ efficacy overall/use in future ^c : "To what degree have you found that including EQ as part of your out-of-class studying has been helpful? Would you continue to use EQ in future courses, even if not formally required to do so?"	Exceptionally	12.0	95.0	45.7
	Very	41.7	86.5	56.9
	Moderately	40.7	86.7	60.2
	Not, Not very	5.6	101.0	45.8

* $p < 0.0005$, one-way analysis of variance (ANOVA); † $p < 0.001$, one-way ANOVA; ‡ $p < 0.03$ nonpaired two-tailed t test; § $p < 0.01$ nonpaired two-tailed t test.

Final exam scores and overall class rank (Biology 360, fall 2007, 122 total students) were tabulated for individual students as a function of their perception of the usefulness of EQ as a study strategy and a measure of the quality of EQ sessions.

^a Survey responses at week 6 of the semester (112 responders).

^b The survey task at week 6 of the semester was, "Write down the one question that was posed to you during one of your EQ sessions that you believe has been the most beneficial in helping you to correct a misconception that you had been harboring up to that time."

^c Survey responses at week 12 of the semester (109 responders).

sions. No attempt was made to correlate that participation with performance on exams.

DISCUSSION

Empirical Evidence of the Utility of Cloning Strategies

Too often, products and methods alleged to improve education are not assessed in a way that generates convincing evidence; potential users are not persuaded that they should take the leap and implement the suggested innovations. The data we have reported here that assess the efficacy of directed formative assessment and elaborative questioning pedagogies are heartening and justify for us the considerable time and effort that was required to experiment with these course modifications. They are generally applicable to both introductory and advanced courses.

The magnitude of performance improvement during the courses was large. As reported in Table 4, the gain achieved by beginning students on a number of rigorous test items that spanned several levels of cognitive taxonomy exceeded 50%. Moreover, final exam scores improved when formative assessment was included in the course design (Table 5). Similarly, we demonstrated marked improvement in the ability of more experienced cell biology students to analyze experimental results and articulate valid conclusions justified by the data, a difficult task that is not intuitive for most people. This improvement was at least 10% when different students were assessed on the same exam problems across multiple semesters (Figure 3) and ranged from 14 to 63% when the subjects were monitored in a pre-post format (Table 6).

Notwithstanding these positive results, sorting out the causal factors is challenging. Responsible instructors con-

stantly, even subconsciously, work to improve every element of the class from semester to semester. The emphasis is on achieving optimum learning, not, from a research standpoint, on carefully controlling the responsible variables. For example, although Table 5 strongly supports the innovations discussed in this article, other subtle changes undoubtedly contributed to the positive outcome. An additional factor complicating interpretations is that scholastic performance may be subject to a "ceiling effect," such that improvement becomes limited by nonacademic factors (personal circumstances in students' lives) over which teachers have no control.

These measures of improved performance were complemented by positive results from assessment of student affect (Tables 7–9). For example, the data reported in Table 8 indicate that students applaud efforts to help them think analytically; to articulate what they are learning verbally, visually, and in writing; and to engage in a metacognitive examination of their study methods and problem-solving strategies. In these respects, the cloning efforts have succeeded.

Deficiencies in Student Study Strategies

Presumably, these cloning pedagogies succeed because they help correct deficiencies in the study practices of inexperienced students. Survey data from alumni at our institution indicate that across all departments, 18% never studied with others, and an additional 33% did so only rarely, probably the night before exams (Alumni Questionnaire, Office of Institutional Assessment and Analysis, Brigham Young University). Lack of peer study support was mentioned by 17% of undergraduates who abandoned a major in science, math, or engineering (Seymour and Hewitt, 1997). Light (2001) observes that always studying alone "is a particular study habit shared by almost all students who are struggling aca-

Table 8. Affective responses (percentages) to course modifications

Affective response		Biology 360	Biology 120
General class attitudes	Worth the time and effort		
	When you consider what you learned in this course (new knowledge, skills, attitudes, and perspectives, etc.) in light of what it cost in terms of personal study time and effort, was what you learned worth the effort you invested?	n = 224	n = 138
	Definitely yes	50.9	59.4
	Probably yes	36.5	29.0
	Probably not	3.5	7.2
	Definitely not	3.0	1.4
	I'm not sure	6.1	2.9
	Metacognitive emphasis on learning strategies		
	To what extent have the efforts made in the course to promote metacognition (self-analysis of one's habits of thinking) caused you to evaluate your strengths and weaknesses in learning and problem solving and attempt to make improvements?	n = 243	
	Exceptionally helpful	21.3	
Elaborative questioning	Very helpful	34.2	
	Moderately helpful	36.1	
	Not very helpful	8.3	
	Not helpful at all	0	
	Perceived benefit of EQ		
	Which of the following best describes the benefit you derive from holding a weekly EQ session?	n = 139	n = 138
	Nearly always significant/very helpful	26	43.5
	Mostly beneficial/somewhat helpful	41	42.8
	Sometimes beneficial	22	
	Minimally beneficial/not very helpful	6	9.4
Formative assessment	Didn't participate/waste of time	4	2.9
		Scale 0–5*	
	Motivation for and conduct of EQ session	n = 120	
	It's a requirement	2.7	
	To understand fundamental concepts, not just memorize facts	3.4	
	All group members participate	4.4	
	Perceived helpfulness in improving data analysis skill		
	To what extent have the formative assessments that we conducted in the course this semester helped you improve your skill in data analysis?	n = 248	
	Exceptionally helpful	13.7	
	Very helpful	41.1	
	Moderately helpful	41.1	
	Not very helpful	2.8	
	Not helpful at all	1.2	

Data (all as percentages) obtained from course evaluation questionnaires administered mid-semester (Biology 360, winter and fall 2007) or at the conclusion of the course (Biology 120, winter 2006).

* Likert scale, where 0 indicates low motivation and 5 indicates high motivation.

demically." Others, however, report that it is the group of high achievers who are reluctant to study with others, due, in part, to the perception of unequal effort ("social loafing") by some members of study groups (Michaelsen *et al.*, 2004). The weakness of this isolation is the absence of feedback. There is no test of whether learning is occurring.

Not only is study in isolation ineffective because of lack of productive interactions, it often takes a superficial form. For example, "Jack and Jill" have to take an exam tomorrow, so how do they prepare? They are accustomed to rereading the text, rereading the notes they made in class (the definition of osmosis is in the upper right-hand corner of page 4), and perhaps run one more time through their set of flash cards on which they have drawn the biochemical formulas of the amino acids. Sorry to disappoint, Jack and Jill, but the ben-

efits of this kind of study are likely to be fleeting. Data from a recent study by Karpicke and Roediger (2008), and earlier from work by Glover (1989), support the conclusion that testing, not further review will best enhance recall. Most students, however, are not accustomed to engage in an effective pre-exam assessment of their understanding.

The effect of these deficiencies is that students often fail to master the material in a course. For example, consider the estimates students make of how much and how long they remember about university courses. The data presented in Table 10 will probably not come as a surprise. Fifty percent retention after 6 months suggests a rather steep decay curve, but in fact this is probably an overly optimistic estimate. Gardiner (1994) cites several published studies documenting poor retention levels of course material by college students.

Table 9. Perceived benefit of pre-exam study strategies^a

Strategy	% yes	% no
Reread the assigned pages from the text on each topic	42	58
Reread the notes I made in class for each topic	95	5
Participate in group study with my EQ partners	95	5
Participate in group study with people other than my EQ partners	52	48
Practice drawing pictures that illustrate my understanding of the major concepts we've studied in the course up to this point	80	20
Review my performance on the four formative assessment exercises	72	28
Review experimental protocols (assays) and practice articulating the experimental questions being asked in an experiment	97	3

^a Survey results: Biology 360, fall 2007, class period 13 of 42, $n = 120$. "Which of the following activities are you planning to engage in by way of preparation for the first mid-term exam?"

In one study, there was only 40% recall of facts at the end of that lecture, and this dropped to 20% 1 wk later. At the end of a two-semester introductory economics course, sophomores scored <20% higher than those in a control group who did not take the course. They subsequently scored only 10% higher as alumni. Even when students retain information, what they retain is frequently laden with misconceptions, even though reasonable explanations have been provided. Some biological examples include the sources and fate of gases during photosynthesis and respiration (Ebert-May and Lim, 2003), basic physiological processes (Modell *et al.*, 2005), the operations of natural selection (Anderson *et al.*, 2002) or, in our experience, the template-dependent synthesis of macromolecules. Not only are they unaware of these mistakes in understanding but also even when corrected, they tended to retain the old erroneous notions (Macbeth, 2000).

Why EQ and Formative Assessment Help

Our experience persuades us about the fundamental importance of student-to-student interaction (EQ) as a means of

Table 10. Students' perception of their retention of university course material: "What percent of the information you have learned in your university courses do you still remember 6 months after those courses are finished? Choose the one closest estimate"

	% information retained	% of student responses
A	100	0
B	90	0.6
C	75	19
D	50	53
E	25	28
F	10	0

Survey results: $n = 170$ cell biology course students, winter 2007.

cloning the professor. Light (2001) seems to agree. His analysis leads to the conclusion that "whether or not students work together in small study groups outside of class is the single best predictor of how many classes in science they will take." Moreover, the data reported in Table 7 suggest that positive attitudes about EQ and the ability to construct high-quality questions in EQ sessions are strong predictors of superior performance on exams. Smith *et al.* (2009) have recently provided evidence that in-class peer discussion leads to improved performance, even when none of the student participants originally knows the correct answer to the problem under consideration.

When it works as it should, EQ is fundamentally different from what occurs in the conventional study group before an exam. The purpose is not to rehearse the capture of textbook trivia ("Jack, name all the major classes of . . ." or "Jill, how old was Darwin when . . .?"). This is because misconceptions can only be abandoned when a skilled interrogator presents to a student a question that cannot be answered correctly based on false ideas ("If that's true, then why . . .?"). Consider instead this example of an effective EQ question, "Why does tryptophan have both a repressor and attenuator function in bacteria?" In a first attempt, the EQ partner may only attempt recall of the many details. Through multiple iterations, the conversation must eventually evolve to the point that the more complex questions of why or how are addressed. By the time this exercise is successfully completed, rapid dissipation of the magnitude suggested by Table 10 no longer happens. Thus, as expressed by Emerson in his famous address to the Harvard Divinity School faculty in 1838: "Truly speaking, it is not instruction, but provocation, that I can receive from another soul." As applied to the science classroom the principle is that comprehension is not transferable; nothing is truly learned until the student engages in an independent construction process that results in personal assimilation.

An added benefit of EQ is that it fosters an appreciation for and excitement about the grandeur of biology. Moreover, it promotes a life-long scholastic disposition like that suggested in the last two rows of the "Student needs" column of Table 2. These are the kinds of skills and attitudes that will carry over to other courses and equip people to deal more effectively with the nonacademic sides of their lives.

What are the benefits of formative assessment? Teachers are likely to agree with the sentiment that an exam should be a learning experience. However, the learning value for students tends to be meager because there is so much emotion associated with being "judged" and "penalized." After the initial shock, both professor and pupil default to an attitude of "for better or worse, what's done is done," and now the focus is on the next exam. Formative assessments, in contrast, permit a student to monitor progress in a safe environment without the stress and anxiety associated with someone passing judgment that prematurely impacts his or her grade in the course. These low-risk, test-like evaluations are intended to inform, to make students' thinking visible to themselves and to their teacher (Huba and Freed, 2000; Handelsman *et al.*, 2007). They provide valuable opportunities for an honest appraisal of the academic status quo and for making specific plans to improve. After a thorough literature review and a careful theoretical consideration, Black and Wiliam (1998) conclude that "Ongoing assess-

ment plays a key role—possibly the most important role—in shaping classroom standards and increasing learning gains,” and “formative assessment . . . is at the heart of effective teaching.”

Anecdotal expressions from students are strongly appreciative of the perceived benefits of a course design with a formative focus. In cell biology, they acknowledge the constructive role of this pedagogy on the development of data analysis proficiency (Table 8). In the Biology 120 course, a correlation analysis showed a stronger relationship between the frequency of formative practice and performance on the final exam than for other measures of effort (number of weekly assessments taken, $r = 0.407$; class attendance, $r = 0.365$; homework completion, $r = 0.261$; completion of text reading assignment, $r = 0.242$; $p < .01$ for all variables). Smith (2007) has also demonstrated a positive impact of a similar mode of formative assessment on exam performance in undergraduate geosciences courses of 25–40 students.

The indispensable element in effective formative assessment is feedback from the teacher, teaching assistants, peers, and one's self. Thus, a productive session will be metacognitive. Students will be provoked to ask themselves the following types of questions: “Why did I answer the way I did?” “Why did I leave out important elements that I had studied?” “What did I do right in preparation that enabled me to answer the problem correctly?” “How can I ensure that I will be able to do that again in the future?” Consider, for example, what might happen during the debriefing that followed an assessment problem of the kind illustrated in Figure 2B. A perplexed student asks, “I thought the idea was that it takes three nucleotides to make one amino acid, but how is this possible if the molecular weight of three nucleotides (310 g/mol each) is approximately 9 times greater than the molecular weight of an amino acid (120 g/mol)?” A classmate sitting in the next seat recognizes the problem. “No,” she suggests, “you don't make amino acids out of nucleotides.” The misconception lies in not understanding the template-dependent nature of translation. “Look at this diagram with me,” she continues, “where are the nucleotides and what are they doing?” The dialogue continues until the first student is able to restate the principle correctly: “It takes three nucleotides in the template mRNA in order to program the insertion of one amino acid in the growing protein.”

The formative scheme explained here owes its effectiveness to two features. The first is the intense effort to analyze performance and make recommendations for improvement. For example, after sessions in which they have been asked to draw valid conclusions from experimental data, students will repeatedly hear these professorial injunctions “Be specific; be thorough; include the obvious; apply before you invent; formulate the experimental question being asked; identify the experimental technique being used and be reminded of the question it is capable of answering; don't just restate the data—interpret their meaning.” It is the extended, reciprocal nature of this conversation that makes it helpful; such an exchange is not normally possible when a student receives a one-time, one-way written comment on a homework problem. The latter certainly has merit; Krasne *et al.* (2006) have provided evidence for performance gains after formative assessment for first-year medical students using only online feedback. However, without the kind of

dialogue illustrated above, students often misunderstand the intent and meaning of the feedback, and the instructor is unlikely to be aware of the students' misunderstanding. Moreover, this scheme makes much more efficient use of a teacher's time; everyone in the class has the opportunity to benefit simultaneously.

The second essential feature is the number of iterations (up to 12 times) of this experience during a semester. This scheduled use of formative assessment has been described as the “embedded-in-the-curriculum” mode, in contrast to its intermittent use, which has been termed “on the fly” (Shavelson *et al.*, 2008). It is probably impossible to overstate the need for repeated practice.

Caveats and How They Might Be Addressed

At some point in the cost–benefit analysis a teacher makes before launching a course reform, he or she should consider the potential negative consequences. Class time will be insufficient for both the traditional lecture and the activities proposed here; these methods require hard work from the instructor, and students and colleagues may complain.

Although obvious, it needs to be emphasized that making changes to a course in an effort to improve teaching and learning demands a compensatory reduction in subject matter coverage; the new cloning strategies cannot be applied as add-ons to business as usual. Allen and Tanner (2007) recognize this need for a “leaner curriculum”: “Who can fail to be aware that the typical life sciences textbook contains too much material for the typical one- or even two-semester course?” They recommend a pruning of subtopics so that there “will be more time left for high-priority learning goals.”

Not only should the curriculum for a course be stripped of excessive, if interesting, detail, but also responsibility for the initial exposure to fundamental concepts must devolve to the students. If you interviewed a typical student at the end of a class period and asked whether he or she had read the assigned pages of the text before coming that day, you would likely hear the following. “I know I'm supposed to read before class, but I wait until sometime afterward. The lecture is easier to understand because my teacher filters out for me what is really important. Otherwise getting from the book what I need for the test is too difficult.” When class time is spent exclusively as a lecture that repeats the content of information contained in the text, a teacher inadvertently helps to perpetuate this abdication of intellectual responsibility.

Students would be much better served if they were provided both instruction and practice in applying proven methods of inquiry to what they read (reading reflectively, formulating appropriate questions, translating ideas into one's own terms, for example; Paul and Elder, 2003). They would learn that the instruction “Please read the assigned pages before coming to class” is not optional because their professor will not take class time to introduce the basics. It would be clear that coming to class unprepared is likely to have a significant negative impact on their performance. Instead of restating information contained in the reading, the professor will use the class period to help students assess their capture of the concepts, and practice their application in a problem-solving setting.

A potential criticism of the formative assessment model is, “But you’re teaching to the test!” Yes, of course we are. Assessment problems do anticipate those on the final exam. However, the exam is the clearest, most direct indicator of what the course objectives are. If you have the right test, teaching to the test is the correct, even mandatory, thing to do. Every football coach teaches to the test as he prepares his players to meet next week’s opponent. Teaching to the test is only bad if the test is superficial and the preparation for it does not provide genuine intellectual growth.

It is our experience that innovations will not meet with unanimous approval. In the courses described here, there is greater emphasis on the development of scientific reasoning skills than on mastery of subject matter detail. Because analytical thinking is difficult and generally underdeveloped, students who can no longer depend completely on memorization are easily frustrated and may express dissatisfaction. Having learned to navigate the traditional system in which they play a primarily passive classroom role, some students object to changes in the routine and prefer to be left alone in their comfortable isolation in the back row. When teachers use the class period to help students assess, articulate, and apply their understanding from assigned reading rather than reiterate the basics from the book, some students complain: “My teacher is not doing his job; he’s not teaching me.” The shift of responsibility to the students for this phase of the learning cycle will be perceived by a few as burdensome. Also, some juniors and seniors often see themselves as beyond the need to experiment with better ways to learn and just want to be left to finish their degrees: “I already know how to learn” is their protest.

Teaching assistants (both graduates and undergraduates) can be extremely helpful in mitigating or eliminating student concerns and assisting with the pedagogies described in this article. Instead of sitting silently in a corner of the classroom, when discussions are taking place they should be roaming the isles, and as members of the academic team, interjecting themselves into conversations, offering encouragement, discussing student worries, recognizing and correcting misconceptions, and asking leading questions to model how a good EQ session operates. They should do the same during the “postmortem” phase of a formative assessment, assisting the members of the class in their attempts at a metacognitive analysis of their performance. Course alumni can also be used in a counseling role, as described in *Materials and Methods*. They provide an excellent perspective on how these practices produce rewards in the end.

Concluding Remarks

There is a clear national concern that many young people lose an interest in science, or at least are deflected from majoring in a scientific subject, while attending college (Seymour and Hewitt, 1997). Our response to this unfortunate trend is that “Students don’t hate biology; they hate the way biology is taught.” The main problem in biology is probably the overemphasis on factual information at the expense of uncovering the beauty of fundamental principles and seeing their meaningful application in people’s personal lives. When a person’s scholastic experience is dominated by passive hours in the classroom and rote memorization of abstractions and interminable lists of unfamiliar vocabulary

terms, the intrinsic capacity of the subject to inspire awe and wonder is diminished. Richard Light suggests a potential remedy with which we agree:

...science professors who succeed in structuring their classes and labs to help undergraduates work collegially are praised by students. The word ‘inspiring’ is often used. These professors attract specialists in both sciences and other disciplines to their classes. Their success is not due to some mysterious charisma, or to their entertainment talents. It is due to the way they organize the work in their courses (Light, 2001).

It is therefore at the level of changes in course management and pedagogy, where cloning efforts are likely to help solve declining student interests.

Part of the reason that a collegial pedagogy succeeds is because of the nurturing it provides, what might be called the “leadership function” of teaching. It fosters the perception that one’s teacher really cares. This is achieved when a teacher leaves the podium, squeezes into the middle of a row, and engages a small group of students face-to-face. It happens when the teacher offers personal advice about how to study more effectively—what to do and not do the night before the final exam, for example. It grows in faculty mentoring sessions, when students receive individualized attention in an informal setting. What these examples share in common is that the psychological distance between teacher and student is reduced. The view that emerges is “My teacher demonstrates an unanticipated and uncommon interest in the success of the students in this class, and is going beyond the call of duty to help people really learn. If my teacher is going to expend this extra effort on my behalf, I’m going to reciprocate in kind.” If we succeed as teachers, we do it one student at a time. When we “clone” those educational functions for which we are best suited and which meet students’ most important needs, the number of those positively affected can be significant.

ACKNOWLEDGMENTS

The contents of this article were developed through U.S. Department of Education (Fund for the Improvement of Postsecondary Education [FIPSE]) grant P116B041238. However, the contents do not necessarily represent the policy or opinion of the Department of Education or the government of the United States.

REFERENCES

- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Molecular Biology of the Cell*, 4th ed., New York: Garland Science.
- Allen, D., and Tanner, K. (2007). Putting the horse back in front of the cart: using visions and decisions about high-quality learning experiences to drive course design. *CBE Life Sci. Educ.* 6, 85–89.
- Anderson, D. L., Fisher, K. M., and Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *J. Res. Sci. Teach.* 39, 952–978.
- Anderson, L. W., and Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom’s Taxonomy of Educational Objectives*, New York: Longman.

- Black, P., and Wiliam, D. (1998). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan* 80, 139–148.
- Bok, D. (2006). *Our Underachieving Colleges*, Princeton, NJ: Princeton University Press.
- Chase, W. G., and Simon, H. A. (1973). Perception in chess. *Cogn. Psychol.* 4, 55–81.
- Ebert-May, D., and Lim, H. (2003). Disciplinary research strategies for assessment of learning. *Bioscience*, 53, 1221–1228.
- Fink, L. D. (2003). *Creating Significant Learning Experiences*, San Francisco, CA: Jossey-Bass.
- Freeman, S. (2004). *Biological Science*, 2nd ed., Upper Saddle River, NJ: Prentice Hall.
- Gardiner, L. F. (1994). Redesigning Higher Education: Producing Dramatic Gains in Student Learning, ASHE-ERIC Higher Education Report, Vol. 23, No. 7, Washington, DC: George Washington University, Graduate School of Education and Human Development.
- Glover, J. S. (1989). The testing phenomenon: not gone but nearly forgotten. *J. Educ. Psychol.* 81, 392–399.
- Goodlad, J. I. (1984). *A Place Called School*, New York: McGraw-Hill Book Company.
- Guyette, W. A., Matusik, R. J., and Rosen, J. M. (1979). Prolactin-mediated transcriptional and post-translational control of casein gene expression. *Cell*, 17, 1013–1023.
- Handelsman, J., Miller, S., and Pfund, C. (2007). *Scientific Teaching*, New York: W.H. Freeman and Company.
- Huba, M. E., and Freed, J. E. (2000). *Learner-centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning*, Needham Heights, MA: Allyn & Bacon.
- Karpicke, J. D., and Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science* 319, 966–968.
- Kitchen, E., Bell, J. D., Reeve, S., Sudweeks, R. R., and Bradshaw, W. S. (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.
- Kitchen, E., King, S. H., Robison, D. F., Sudweeks, R. R., Bradshaw, W. S., and Bell, J. D. (2006). Rethinking exams and letter grades: how much can teachers delegate to students? *CBE Life Sci. Educ.* 5, 270–280.
- Krasne, S., Wimmer, P. F., Relan, A., and Drake, T. A. (2006). Differential effects of two types of formative assessments in predicting performance of first-year medical students. *Adv. Health Sci. Educ.* 11, 155–171.
- Light, R. J. (2001). *Making the Most of College: Students Speak Their Minds*, Cambridge, MA: Harvard University Press.
- Macbeth, D. (2000). On an actual apparatus for conceptual change. *Sci. Educ.* 84, 228–264.
- McClymer, J. F., and Knowles, L. Z. (1992). Ersatz learning, inauthentic testing. *J. Excell. Coll. Teach.* 3, 33–50.
- McDaniel, M. A., and Donnelly, C. M. (1996). Learning with analogy and elaborative interrogation. *J. Educ. Psychol.* 88, 508–519.
- Michael, A. L., Klee, T., Bransford, J. D., and Warren, S. (1993). The transition from theory to therapy: test of two instructional methods. *Appl. Cogn. Psychol.* 7, 139–154.
- Michaelsen, L. K., Knight, A. B., and Fink, L. D. (2004). *Team-Based Learning. A Transformative Use of Small Groups in College Teaching*, Sterling, VA: Stylus Publishing.
- Modell, H., Michael, J., and Wenderoth, M. P. (2005). Helping the learner to learn: the role of uncovering misconceptions. *Am. Biol. Teach.* 67, 20–26.
- Overtom, C. (2000). Employability skills: an update. *Eric Digest* No. 220. http://brownfields-toolbox.org/pdf_toolbox/employability220.pdf (accessed 7 March 2009).
- Paul, R., and Elder, L. (2003). *How to Read a Paragraph and Beyond. the Art of Close Reading. How to Read a Text Worth Reading and Take Ownership of Its Important Ideas*, Dillon Beach, CA: The Foundation for Critical Thinking.
- Roselli, C. E., Larkin, K., Resko, J. A., Stellflug, J. N., and Stormshak, F. (2004). The volume of a sexually dimorphic nucleus in the ovine medial preoptic area/anterior hypothalamus varies with sexual partner preference. *Endocrinology* 145, 478–483.
- Seymour, E., and Hewitt, N. M. (1997). *Talking about Leaving: Why Undergraduates Leave the Sciences*, Boulder, CO: Westview Press.
- Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., Tomita, M. K., and Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: a collaboration between curriculum and assessment developers. *Appl. Meas. Educ.* 21, 295–314.
- Smith, G. (2007). How does student performance on formative assessments relate to learning assessed by exams? *J. Coll. Sci. Teach.* 36, 28–34.
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., and Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science* 323, 122–124.
- Tanner, K. D. (2009). Talking to learn: why biology students should be talking in classrooms and how to make it happen. *CBE Life Sci. Educ.* 8, 89–94.