Biology Majors' Knowledge and Misconceptions of Natural Selection

ROSS H. NEHM AND LEAH REILLY

This article reports on a study that assessed knowledge of and misconceptions about natural selection in second-semester biology majors in two classes characterized by different instructional strategies. The active-learning class achieved significant postcourse gains in the number and diversity of key concepts of natural selection employed in evolutionary explanations and exhibited significant decreases in misconception use. Compared with the traditionally taught class, the active-learning class was characterized by fewer misconceptions and greater mean key-concept diversity scores. Nevertheless, both classes demonstrated inadequate postcourse levels of evolutionary understanding: After a year of college biology, 70 percent of students in the active-learning group and 86 percent in the traditionally taught group employed one or more misconceptions in their evolutionary explanations. Faculty in upper-division courses must be prepared to address students' misconceptions and provide additional opportunities for improving student understanding of natural selection.

Keywords: evolution, education, misconceptions, natural selection

Low levels of evolutionary knowledge and high levels of evolutionary misconceptions are known to be harbored by high school students (Clough and Wood-Robinson 1985, Demastes et al. 1995), undergraduates (Bishop and Anderson 1990), biology majors (Dagher and BouJaoude 1997), medical students (Brumby 1984), and science teachers (Affanato 1986, Osif 1997, Nehm and Schonfeld forthcoming). Considering that low levels of evolutionary knowledge are pervasive in many of the groups that have taken introductory biology, and that introductory biology is one of the most highly enrolled undergraduate science courses in the United States, one might ask what impact, if any, such courses are having on students' knowledge and misconceptions of natural selection.

Many studies have focused on nonmajors' and first-semester biology students' knowledge, antievolutionary attitudes, and misconceptions (Brumby 1984, Bishop and Anderson 1990, Jensen and Finley 1997). Unfortunately, these studies tell us very little about what intended majors know about natural selection as they move on to advanced biology coursework, because many of the students who enroll in first-semester biology do so for the sole purpose of satisfying their laboratory science requirement, and they do not intend to major in biology. Therefore, a large gap in the evolution education research literature exists between first-semester biology students and nonmajors at one end of the spectrum (Bishop and Anderson 1990) and medical students and biology teachers at the other (Nehm and Schonfeld forthcoming). The goal of our study is to begin to fill this gap by exploring the evolutionary knowledge and misconceptions of second-semester biology students—that is, those students pursuing a major in biology.

This study addresses three questions: (1) What magnitude of knowledge and of misconceptions about natural selection do second-semester biology majors bring to the classroom? (2) What is the magnitude of the knowledge of natural selection gained during a course taught using an active-learning approach? (3) What levels of knowledge and misconceptions of natural selection characterize biology majors after a year of instruction?

Sample characteristics

This study was conducted on two classes of second-semester biology majors at a college located in an urban area of the northeastern United States. In one of the classes, the instructor employed an active-learning teaching strategy, with

Ross H. Nehm (e-mail: rnehm@ccny.cuny.edu) is an associate professor in the Department of Biology and the School of Education, The City College, City University of New York (CUNY), New York, NY 10031. Leah Reilly (e-mail: leah.reilly@gmail.com) is a graduate student in the Ecology, Evolutionary Biology, and Behavior program at CUNY.

evolution as a common thread through all units; in the second, the instructor employed a more traditional lecturing style, with one discrete unit on evolution. Enrollment in the courses was contingent on successful completion of the first semester of introductory biology, which covered genetics and cell biology. The mean age of students in both groups was 21 years (range: 17 to 36 years). Females comprised 61 percent of both groups. The racial and ethnic distribution of the students in the study closely approximated that of all science students at the institution (Hispanic, 32.5 percent; African American, 30.12 percent; Asian, 25.5 percent; American Indian, 0.09 percent; and white [non-Hispanic], 11.75 percent), although the traditional pedagogy group had a slightly higher percentage of white non-Hispanic students (15 percent versus 10 percent). Eighty-two students from the active-learning group participated in the survey (82 percent response rate), as did 100 students from the traditional-learning group (99 percent response rate). Overall, our sample is different from almost all previous evolution education studies in that it comprises mostly minority undergraduate biology majors, slightly older students, and a greater proportion of females.

The course interventions: Curricula and pedagogies

The 12-week, second-semester introductory biology course that served as the "treatment" group for this study covered standard topics in organismal biology but employed evolution as a unifying theme (table 1). Although student misconceptions have been documented in many areas of biology (Duit 2006), a primary emphasis in the intervention was to increase students' working knowledge of natural selection and reduce their misconceptions of it. Although content coverage is a serious problem in rapidly evolving fields like biology—many biological topics are important for students to learn—emphasis on natural selection in introductory biology does not greatly diminish content coverage of other areas.

The pedagogical component of the intervention was based on an active-learning model. Considerable evidence suggests that active-learning strategies in general, and cooperative learning in particular, enhance student engagement and performance (NRC 2000, 2003). Cooperative learning also downplays competitive or individualistic aspects of science learning that often alienate students, particularly those from underrepresented groups. The course also abandoned exclusively lecturing and administering tests that rewarded factual recall, because these approaches are at odds with a large body of research on how people actually learn (NRC 2000, 2003). In sum, students learn best through active participation in the learning process (NRC 2000), and therefore our experimental course employed numerous opportunities for active learning (table 1).

Many active-learning strategies have been developed for addressing antievolutionism in general and student misconceptions of natural selection in particular. These include inquiry instruction (Demastes et al. 1995), paired problem solving (Jensen and Finley 1997), small-group discussion (Scharmann 1993), historically rich curricula (Jensen and Finley 1997), modeling approaches (Passmore and Stewart 2002), explicit discussion of religious–scientific boundaries (Gould 1999), explorations of the nature of science (Dagher and BouJaoude 1997), and emphasis on formal reasoning and critical thinking skills (Lawson and Worsnop 1992). Our active-learning group was exposed to discussions of the nature of science, paired problem solving, small-group discussions, and group response questions in every class (see table 1).

Both groups had the same instructional time, textbook, reading assignments, and lab experiences, but different instructors. Both instructors considered evolution a unifying course principle. The comparison class experienced lecturing exclusively and was characterized by the absence of active learning (as defined above; table 2). In addition, the comparison group was taught evolution within a discrete unit at the beginning of the course, in contrast to the integrated curriculum of the treatment group. The treatment group was taught using PowerPoint lectures containing numerous images and text, whereas the comparison group was taught using a chalkboard containing handwritten notes, drawings, and diagrams. The treatment group was provided with a photocopy of the PowerPoint slides, and the comparison group was provided with a photocopied set of textbook diagrams to accompany the lecture.

Instrument characteristics

A paper-and-pencil instrument was developed in order to measure student knowledge and misconceptions of natural selection using results from previous research (and a pilot study; Bishop and Anderson 1990, Nehm and Schonfeld forthcoming). Although Anderson and colleagues (2002) developed the multiple-choice Conceptual Inventory of Natural Selection (CINS) to measure similar types of knowledge, their instrument was poorly matched to the setting of our study in several respects. First, the CINS was employed and validated on a cohort of undergraduate nonmajors, whereas we are interested in studying biology majors in their second semester. Second, our second-semester introductory biology students have been demonstrated to perform significantly differently on open-ended than on multiple-choice reading comprehension exams, possibly because a sizable fraction of students are English-language learners. Third, an open-response format is indicated when predicting all important and likely responses in advance is impossible (Ary et al. 2002); moreover, since a main goal of the study was to establish an accurate baseline level of knowledge brought by second-semester undergraduates, the benefit of capturing unanticipated misconceptions outweighed the disadvantages of essay instruments. A forthcoming study, however, will explore the validity of both the CINS and the open-response instrument, as well as the ways in which the instruments differentially evoke student knowledge and misconceptions of natural selection as compared with oral interviews (Nehm and Reilly 2007).

Table 1. Course topics and examples of active-learning activities.

Schedule	Topic	Example of active-learning activity
Week 1	What do biologists do and how do they think about life? (field versus lab research; scales of biological research; experimentation modes; patterns and processes; hypothesis testing; models and modeling)	Group problem solving on the differences between theories and laws in biology
Week 2	What processes produce variation? (mutation, meiosis, recombination, and sex)	Paired problem solving: Can variation be stopped?
Week 3	How is variation transmitted from generation to generation? (Mendelian genetics)	Paired problem solving: Mendelian genetics problems
Week 4	How do genotypes help to build phenotypes? (molecular genetics and development)	Group response questions about mutations
Week 5	Why is life always changing? (part I: microevolutionary and macro-evolutionary patterns, natural selection)	Paired problem solving and group discussion: identifying reproductive barriers to gene flow
Week 6	Why is life always changing? (part II: microevolutionary patterns, natural selection, and genetic drift)	Paired and group discussion: Does selection always occur?
Week 7	Why is life always changing? (part III: species, speciation, phylogenetics, and biodiversity)	Using Genbank and Biology Workbench to develop a molecular phylogeny of bears
Week 8	What is the role of energy in biology? (part I: energetics and biomechanics)	Group response questions: interpreting phylogenetic trees of elephants
Week 9	How do organisms function? (part I: form, function, and energy)	Paired problem solving: calculating flow rate and food capture in a sponge
Week 10	How do organisms function? (part II: feeding, digestion, and osmoregulation)	Paired problem solving: water relations and movement in trees
Week 11	What is the role of energy in biology? (part II: energetics, food webs, and global cycles)	Group problem solving: ranking abundances of marine species on the basis of energetics and food webs
Week 12	How will climate change influence life? (invasive species, extinction, communities, and succession)	Paired problem solving: calculating population size growth of a marine snail with and without resource limitations

Table 2. A comparison of the traditional and active-learning groups involved in a study of two second-semester biology (Biology 102) classes, along with postcourse knowledge measures.

Variable	Traditional lecture with discrete evolution content	Active learning with integrated evolution content
Characteristics of study participants		
Number of students	101	100
Number of voluntary study participants	100	89
Mean age	21	21
Percentage female	61	61
Percentage white (non-Hispanic)	15	10
Percentage who successfully completed Biology 101	100	100
Percentage who had heard about the idea of natural selection	99	99
Percentage who reported being taught natural selection in school	94	84
Characteristics of course		
Percentage of classes involving active learning	0	86
Percentage of classes discussing natural selection content knowledge	20	90
Postcourse knowledge measures		
Percentage of students employing four or more key concepts	58	70
Mean key-concept diversity	3.78	4.33
Percentage of students employing no misconceptions	14	30
Mean misconception diversity	1.91	2.41
Natural selection performance quotient	0.74	0.79

Our open-response paper-and-pencil instrument was designed to be completed during class in 25 minutes or less. In addition to basic demographic variables, we asked students to report whether they had ever heard about the idea of natural selection or had been taught about it in school. We then asked a series of open-ended essay questions used in previous studies (box 1; Bishop and Anderson 1990, Nehm and Schonfeld forthcoming). Students had half a page for answering each question, and they were asked to "be as complete as you can" both on the instrument and in an oral script. While students were working, the proctors again instructed them to answer as completely as possible.

Overall, our instrument questions were designed to determine how successful biology majors are at answering questions about natural selection at differing levels of complexity. The six questions were ordered such that they began by requesting familiar, concrete knowledge (e.g., "define natural selection") and ended with unfamiliar, abstract problem-solving questions (e.g., "If biologists wanted to speed up evolutionary change, how would they do it?"). These questions, all of which focused on natural selection, were designed to span several of Bloom's taxonomic levels, which categorize types of knowledge into hierarchical levels of complexity, from lowest to highest: knowledge, comprehension, application, analysis, synthesis, and evaluation (Bloom 1956).

Box 1. Open-response instrument essay questions.

1. Please define natural selection to the best of your ability.
2. List and describe the six components (or parts) of natural selection.
3. Explain why some bacteria have evolved a resistance to antibiotics (that is, the antibiotics no longer kill the bacteria).
4. Cheetahs (large African cats) are able to run faster than 60 miles per hour when chasing prey. How would a biologist explain how the ability to run fast evolved in cheetahs, assuming their ancestors could run only 20 miles per hour?
5. Cave salamanders (amphibian animals) are blind (they have eyes that are not functional). How would a biologist explain how blind cave salamanders evolved from ancestors that could see?
6. If biologists wanted to speed up evolutionary change, how would they do it?

Note: Question 5 was not part of the precourse instrument.

Bloom's level 1, knowledge, refers to cognitive tasks involving the basic observation and recall of information, such as an awareness of dates, events, places, and major ideas, elicited by prompts such as "define" and "list." In contrast, Bloom's "application" category includes tasks such as employing information, methods, or concepts in new situations to solve problems, in response to requests such as "explain" or "design." Thus, our instrument provided students with multiple opportunities to solve evolutionary problems at several levels of complexity.

Variables

The first set of variables extracted from the instrument related to student knowledge of seven "key concepts" of natural selection (Mayr 1982): (1) the causes of phenotypic variation (e.g., mutation, recombination, sexual reproduction), (2) the heritability of phenotypic variation, (3) the great reproductive potential of individuals, (4) limited resources or carrying capacity, (5) competition or limited survival potential, (6) selective survival based on heritable traits, and (7) a change in the distribution of individuals with certain heritable traits.

A coding rubric was developed, piloted, refined, and used to score student responses, such that the use of a key concept in an explanation of evolutionary change counted as one point. To test the precision of the coding rubric and the consistency of the raters (i.e., interrater reliability, or IRR), after the initial coding the essays were blindly recoded using the same rubric. IRR was measured using Pearson correlation coefficients of key concepts between the two raters. The results indicated statistically significant correlations for the two questions that were examined ($r = 0.784$, $p < 0.001$; $r = 0.768$, $p < 0.001$). Thus, the scoring rubric appeared to be sufficiently clear, and the raters sufficiently consistent, for coding the presence or absence of key concepts in students' essay responses. The coding rubric was used to quantify the presence or absence of the seven key concepts in each of the students' six essay questions. These scores were tallied separately for each question, collectively for each student, and collectively for all students (pre- and postcourse).

The second set of variables extracted from the instrument related to student misconceptions of natural selection. We developed a coding rubric that contained commonly documented misconceptions about natural selection and evolution from the literature (Bishop and Anderson 1990). We used this rubric to score the magnitude and distribution of commonly observed student misconceptions—for example, needs cause evolutionary changes to take place, the use or disuse of traits explains their appearance or disappearance, traits appear only when they are needed, all individuals in a population develop new traits simultaneously (Bishop and Anderson 1990, Nehm and Schonfeld forthcoming)—and to capture any novel misconceptions elicited by the instrument.

Student responses were scored such that the use of an identifiable misconception in an evolutionary explanation counted as one point, with no upper limit on the number of misconceptions recognized per essay. Unanticipated misconceptions captured in this manner included the belief that (a) "survival of the fittest" means survival of the fittest *species;* (b) "fit" means dominant and "unfit" recessive, in the allelic sense; (c) "genetic drift" is gene flow between different species; (d) drastic climate change is required for evolution to occur; and (e) heritable "compensation" of one trait occurs when another faculty is lost (e.g., "super" hearing or smell was attributed to suddenly blind salamanders). The coding rubric was used to score the presence or absence of these misconceptions, and scores were tallied for each question, collectively for each student, and collectively for all students (pre- and postcourse).

In addition to studying student performance using separate measures of the abundance and diversity of key concepts and of misconceptions, we developed a single measure—the natural selection performance quotient (NSPQ)—to quantify student knowledge and misconceptions. The NSPQ takes a ratio of key-concept diversity to the sum of key-concept diversity and misconception diversity, multiplies it by the ratio of key-concept diversity to total possible key concepts, and produces a single performance score on a 0 to 100, gradelike scale. The first term expresses the proportion of the students' answers that were correct, and the second expresses how the correct proportion compared to the most complete possible answer. Exponents were chosen to calibrate the NSPQ scale so that it conformed to our assessment that four key concepts would result in a score greater than 65. In addition to permitting the visualization of student knowledge on a single scale, the NSPQ also distinguishes clearly between students who have problems with their understanding of natural selection, despite displaying significant knowledge, and those students with no misconceptions who displayed differing levels of knowledge (table 3).

Results

Biology majors in the beginning of their second semester reported hearing of and being taught about the idea of natural selection before enrolling in the course. Ninety-nine percent of students in both the comparison and the treatment groups reported having heard about the idea of natural selection. Ninety-four percent of the comparison group respondents and 83.8 percent of active-learning group respondents reported being taught about the idea in school. However, in their precourse definitions of natural selection, only 3.2 percent of the active-learning group of students employed four or more key concepts, which we consider to be an adequate approximation of understanding. In their definitions of natural selection, 27.4 percent of students did not mention a single key concept, 37.9 percent of students mentioned one key concept, 22.1 percent mentioned two key concepts, and 9.5 percent mentioned three key concepts. In addition to their incomplete definitions of natural selection, misconceptions were present in 29.5 percent of student definitions. Thus, despite having heard of and having been taught about natural selection, very few students could provide an accurate definition of it as measured by the precourse instrument. In addition, a large number of students began the second-semester class harboring misconceptions about natural selection and evolution.

Precourse, in each of the five essay questions, students employed an average of about one key concept. Postcourse, mean key-concept scores increased significantly for all five questions in the active-learning group (figure 1, table 4). Although statistically significant gains were made in the number of key concepts employed by students in their explanations of evolutionary change, most students' postcourse repertoire of key concepts still fell short of constituting an adequate definition of the theory of natural selection (figure 1).

Across the six postcourse essay responses, we coded the number of different key concepts used by each student (the students' key-concept diversity score). A score of 7 was the highest possible diversity score, meaning that the individual student employed all seven key concepts of natural selection at some point on the instrument. Because of the overlapping content of the instrument questions, we argue that the diversity scores are the most meaningful indicators of total student knowledge and misconceptions; individual question scores and their respective elicitation of key concepts are more informative

Table 3. Natural selection performance quotient scores.

Key concepts	Misconceptions	NSPQ
0	0	0.00
0	1	0.00
0	2	0.00
0	3	0.00
0	4	0.00
0	5	0.00
0	6	0.00
1	0	0.70
1	1	0.58
1	2	0.51
1	3	0.47
1	4	0.43
1	5	0.41
1	6	0.39
2	0	0.80
2	1	0.70
2	2	0.64
2	3	0.59
2	4	0.56
2	5	0.53
2	6	0.50
3	0	0.86
3	1	0.78
3	2	0.72
3	3	0.68
3	4	0.64
3	5	0.61
3	6	0.59
4	0	0.90
4	1	0.84
4	2	0.79
4	3	0.74
4	4	0.71
4	5	0.68
4	6	0.65
5	0	0.94
5	1	0.88
5	2	0.83
5	3	0.80
5	4	0.76
5	5	0.73
5	6	0.71
6	0	0.97
6	1	0.92
6	2	0.88
6	3	0.84
6	4	0.81
6	5	0.78
6	6	0.75
7	0	1.00
7	1	0.95
7	2	0.91
7	3	0.88
7	4	0.84
7	5	0.82
7	6	0.79

Note: The natural selection performance quotient distinguishes students who have problems with their understanding of natural selection, despite displaying significant knowledge, from those students with no misconceptions who display differing levels of knowledge. All students who employ more than four key concepts, regardless of their misconception magnitudes, receive a "passing" score of 0.65.

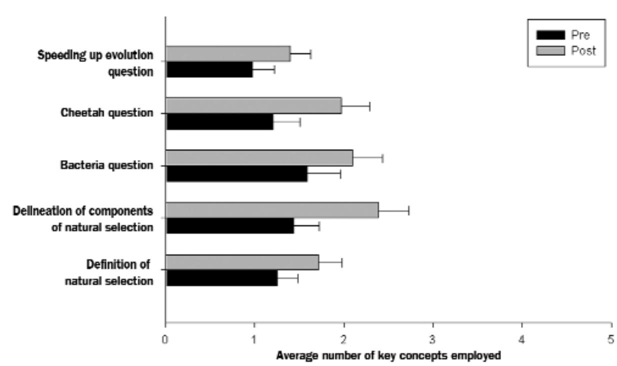

Figure 1. Frequency of key concepts per question pre- and postcourse in the treatment group. See box 1 for a list of the full questions. Error bars represent 2 standard errors about the mean.

Table 4. Pre- and postcourse t *tests for changes in the frequency of key-concept and misconception use by students in the treatment group.*

Category	Question	*t*	df	Significance (two-tailed)
Key concepts	Question 1 (define natural selection)	–2.68	167.8	**0.008**
Key concepts	Question 2 (describe components of natural selection)	–4.26	144.1	**0.000**
Key concepts	Question 3 (explain evolution of bacterial resistance to antibiotics)	–2.03	163.0	**0.044**
Key concepts	Question 4 (explain evolution of cheetah's running speed)	–3.48	162.8	**0.001**
Key concepts	Question 6 (explain how biologists could speed up evolutionary change)	–2.53	143.9	**0.012**
Misconceptions	Question 1	2.11	171.6	**0.037**
Misconceptions	Question 2	2.78	103.5	**0.007**
Misconceptions	Question 3	3.86	147.3	**0.000**
Misconceptions	Question 4	1.82	157.5	0.070
Misconceptions	Question 6	–0.32	144.1	0.749

df, degrees of freedom.

Note: Values in boldface are significant at $p < 0.05$. Questions are listed in full in box 1; question 5 (explain evolution of blind cave salamanders) was not part of the precourse instrument and thus is not included in this table.

about the design of the instrument and the relative usefulness of its parts.

Precourse, the mean diversity of key concepts employed by the active-learning group was 2.8 (standard deviation [SD] 1.8). Precourse, 32.3 percent of students employed no key concepts or one key concept, 28.2 percent employed two or three key concepts, and 39.5 percent employed four or more key concepts. A significant increase in the mean diversity of key concepts used by the active-learning group occurred postcourse (postcourse mean = 4.33; $t = –5.89$, 179 degrees of freedom [df], $p < 0.001$). Postcourse, no students employed zero key concepts (compared with 11 percent precourse). In addition, a much higher proportion of students used four or more key concepts postcourse (69.5 percent, compared with 39.5 percent precourse; figure 2). Postcourse, 58 percent of students in the comparison group employed four or more key concepts (compared with 69.5 percent in the treatment group; table 2). Mean key-concept diversity was also lower in the comparison group (3.78) than in the active-learning group (4.33), although this difference was not statistically significant.

The mean NSPQ increased significantly from pre- to postcourse for the active-learning group (62 to 79; figure 3), and the proportion of students with an NSPQ above 65 increased from 56 percent to 85 percent. The mean NSPQ did not differ significantly between the postcourse traditional and active-learning groups (74 and 79, respectively), but the proportion of students with an NSPQ above 65 (i.e., "passing") did: 74 percent for the traditional and 85 percent for the active-learning group.

Finally, to determine whether the motivational context of instrument implementation could be underestimating students' use of key concepts, we gave students another opportunity to answer a single question—the final, "speeding up evolution" prompt from the pre- and postcourse instruments—for extra credit on the final exam, one week after the postcourse instrument was administered; students did not know in advance that this opportunity would be offered. Comparing the students' use of key concepts on the

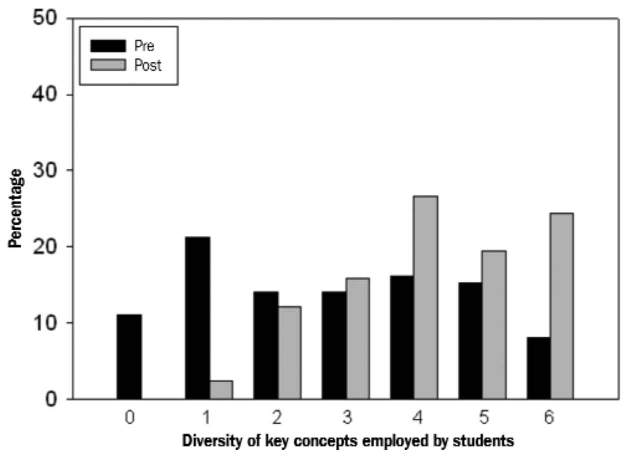

Figure 2. Student composite responses to questions about natural selection, pre- and postcourse, in the treatment group. A key-concept score of 7 was the highest possible diversity score, meaning that the individual student employed all seven key concepts of natural selection at some point in his or her responses. Note the shift in distribution to the right postcourse; a greater percentage of students used a greater diversity of key concepts in their explanations of natural selection.

postcourse instrument's "speeding up evolution" question with their use on the extra-credit essay on the final exam, we found that the postcourse instrument contained on average 0.5 fewer key concepts than did the extra-credit essay (figure 6). Thus it appears that the voluntary context of the postcourse instrument only slightly underestimated student key concepts.

Misconception magnitudes

We explored student composite responses to the instrument questions about natural selection to determine whether the diversity of misconceptions changed from pre- to postcourse in the treatment group. A misconceptions score of 6 was the highest possible diversity score in this measure. Precourse, the mean diversity of misconceptions employed by the active-learning group was 2.47 (SD 2.0). Precourse, 40.4 percent of the students employed no misconceptions or one misconception, 17.2 percent employed two or three misconceptions, and 42.4 percent employed four or more misconceptions.

Although the *frequency* of misconceptions decreased significantly after the course (figure 4), there was no significant change in the *diversity* of misconceptions students employed ($t = 0.198$, 172 df, $p = 0.843$; figure 5). Postcourse, similar percentages of students employed no misconceptions (30.3 percent precourse versus 30.5 percent postcourse), whereas a slight decrease in the percentage of students using four or more different misconceptions occurred (42.4 percent precourse versus 35.4 percent postcourse). In the traditional-learning group, 14 percent of students postcourse had no misconceptions (compared with 30 percent in the treatment group). On average, however, students in the active-learning group who had misconceptions displayed misconception magnitudes (2.41) similar to those of students in the traditional-learning group (1.91).

To determine whether the context of instrument implementation was accurately capturing students' use of misconceptions, we compared student misconception magnitudes between the postcourse "speeding up evolution" question and the extra-credit opportunity discussed previously. We found that the postcourse and extra-credit instruments captured comparable magnitudes of misconceptions (figure 6). Thus our instrument provided a reliable estimate of student misconceptions in different motivational contexts.

Discussion

Considerable research has demonstrated the limited effects of lecture-style pedagogy on student learning (NRC 2000, 2003) and the postinstructional persistence of evolutionary misconceptions (Bishop and Anderson 1990, Nehm and

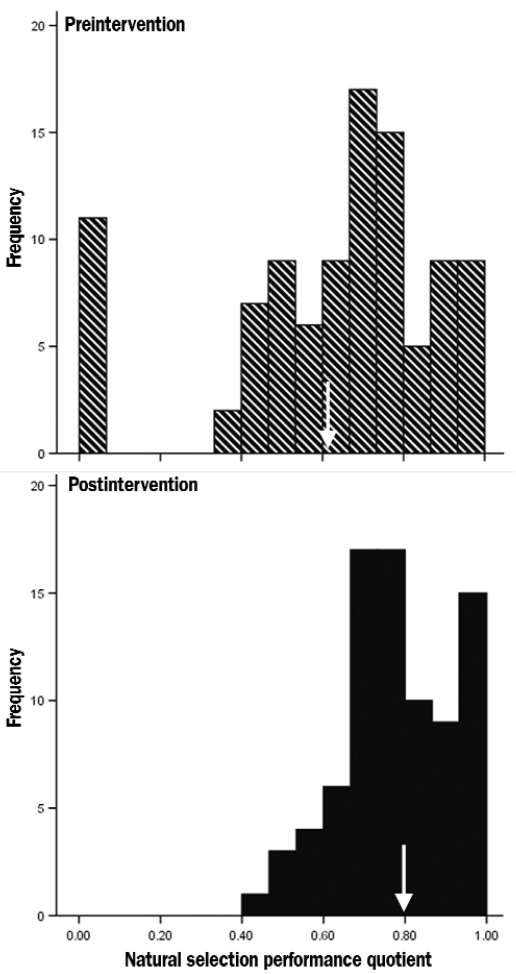

Figure 3. Distribution of scores on the natural selection performance quotient (NSPQ) in the treatment group pre- and postintervention. The NSPQ is a composite measure of the number of key concepts and misconceptions that students employed in their explanations. The arrows illustrate the average score for each group.

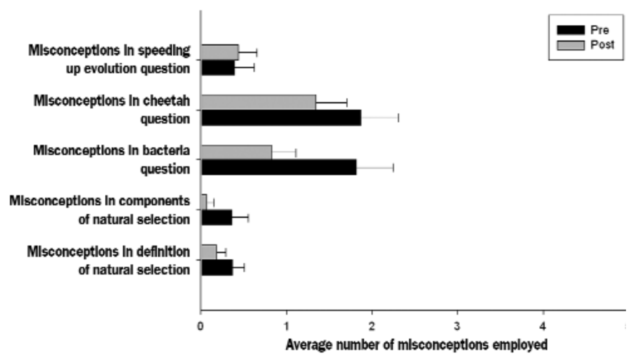

Figure 4. Frequency of misconceptions per question pre- and postcourse in the treatment group. See box 1 for a list of the full questions. Error bars represent 2 standard errors about the mean.

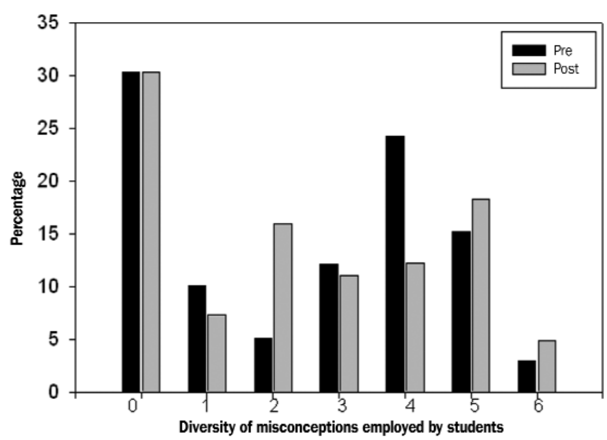

Figure 5. Student composite responses to questions about natural selection in the treatment group. A misconception score of 6 was the highest possible diversity score, meaning that the individual student employed all six misconceptions of natural selection at some point in his or her responses. Note that there is no major change in the distribution pre- to postcourse.

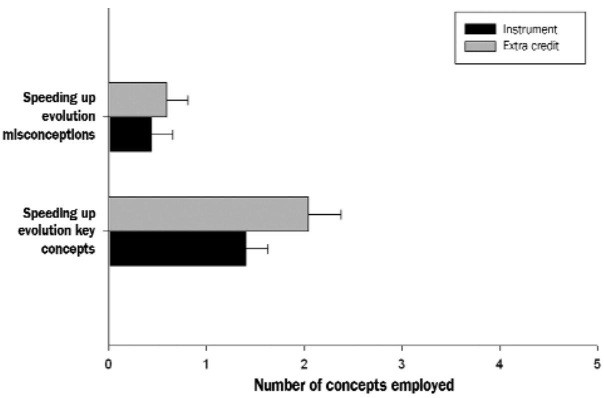

Figure 6. Frequency of misconceptions and key concepts for the "speeding up evolution" question (box 1) under two different contexts (instrument postcourse and a final exam extra-credit question one week later). Error bars represent 2 standard errors about the mean.

Schonfeld forthcoming). As a result, we designed a course intervention infused with active learning and integrated evolutionary content. If prior research serves as an accurate guide, traditional pedagogies coupled with discrete rather than integrated evolutionary content should produce smaller learning gains.

We estimated the magnitudes of correct and incorrect ideas about natural selection held by students using key-concept and misconception scores coded from open-response essay questions. Open-ended questions are best suited to research designs in which all possible and important likely answers cannot be specified in advance (Ary et al. 2002). Open-response questions are also considered to have advantages over closed-response (e.g., multiple-choice) questions in that they provide greater insight into how students think and what they really know (Ary et al. 2002). These advantages were considered to outweigh the disadvantages, which included mismatched responses (e.g., students may not understand what components of their existing knowledge the instrument is prompting, thus producing misleading data) and difficulty in interpreting students' intended meanings. We attempted to compensate for these disadvantages in two ways: (1) Content experts and students read the questions, and any ambiguous or misleading phrases were altered before administering the instrument; (2) we carefully constructed a scoring rubric and quantitatively studied IRR.

Our inability to employ a more rigorous experimental design (e.g., individuals randomly placed into sections and taught identical content by the same instructor, with precourse data for the traditional-learning group) precludes any definitive judgments about the superiority of our active-learning intervention. The active-learning class, however, was characterized postcourse by (a) a greater percentage of students employing more than four key concepts (70 percent versus 58 percent), (b) a higher mean key-concept diversity score (4.33 versus 3.78), (c) a greater percentage of students lacking misconceptions (30 percent versus 14 percent), and (d) greater average NSPQ scores (0.79 versus 0.74).

Despite oral and written prompts to be as complete as possible in their answers, and a half-page of empty space in which to answer each instrument question, it is possible that a lack of motivation to respond fully, rather than a lack of knowledge, led to a misleadingly low diversity of key concepts. We explored this possibility in three ways. First, as noted above, we provided a second opportunity for students to answer a single question from the instrument—the final, "speeding up evolution" prompt—as extra credit on the final exam. If the voluntary context of instrument implementation led to low student effort and subsequently low key-concept scores, then the significantly greater motivational context of extra credit on a final exam should lead to more key concepts and fewer misconceptions. However, we found that the postcourse instrument contained on average only 0.5 fewer key concepts than the extra-credit essay and the same number of misconceptions (figure 6). This suggests that the motivational context of instrument implementation does not fully explain the low magnitudes of key concepts documented in the traditional and active-learning classes.

The second approach that we used to explore the possibility of key-concept underestimation as a result of low student effort was to consider misconception magnitudes. The high levels of misconceptions that we document in both the active-learning and the traditional-learning classes support the conclusion that low student effort alone does not explain the low key-concept scores. If students had a good understanding of natural selection but were unmotivated to use key concepts in their explanations, it is unlikely that they would have the motivation to use numerous misconceptions.

The students' NSPQ scores also indicate that it is unlikely that low student effort explains the low magnitude of key concepts in both the active-learning and the traditional-learning classes. NSPQ scores quantify how many students used only key concepts and no misconceptions in their responses across all six instrument prompts. This student group is the only likely source of "low-effort" respondents who may in fact have harbored significantly more knowledge. However, this group comprises a very small percentage of our samples (less than 16 percent in both pre- and postcourse analyses). Thus, low effort but high knowledge cannot explain the overall patterns that we document.

The third approach examined patterns of nonresponse in the pre- and postcourse active-learning groups. Although our participant response rate on the instrument as a whole was very high (see above), the number of students completing particular questions varied widely. Interpreting the meaning of the response rate to instrument questions is complex (Ary et al. 2002). A lack of student response to a particular question may be due to a lack of knowledge, an inability to employ extant knowledge in the context of the assessment, a lack of clarity in the instrument itself, or combinations thereof.

In designing the instrument, each of the five precourse questions was assigned its place in sequence according to ostensible level of complexity or difficulty (Bloom 1956). The number of people answering each question did decline in roughly their order of occurrence (excepting the "components" question), but is this pattern necessarily attributable to increasing difficulty of the questions? If students with four or more key concepts (a satisfactory grasp of the idea of natural selection, by our standard) left "blanks" at comparable rates to students with low key-concept scores, we would be unable to reject the hypothesis that blanks were due to students' low effort. However, of the 39 students in the active-learning group who scored four or more key concepts at the beginning of the semester, only 8 skipped questions (mean number of questions skipped = 0.26 per student), whereas out of the 60 students with fewer than four key concepts, 28 failed to answer all questions (mean number of questions skipped = 2.0 per student).

This pattern implies (a) that the presence of blanks does, in fact, most likely indicate poor understanding; and (b) that, since a far greater number of students left blanks on the precourse instrument (129 total blanks left by 39 individuals, compared with 20 total blanks left by only 17 individuals postcourse), the precourse misconception scores may underestimate the number of student misconceptions. Thus, the postcourse losses in misconceptions documented in this study may be biased by disproportionate precourse nonresponses. In summary, several different lines of reasoning lead us to conclude that the scores generated from the instrument responses serve as accurate estimates of students' knowledge and misconceptions of natural selection.

Although all of the undergraduates who participated in our study had successfully completed one semester of introductory biology, and most (approximately 83 percent) reported that they had been taught about natural selection in high school, they arrived in second-semester biology with remarkably limited knowledge of natural selection. Only 3.2 percent of students employed four or more key concepts in their explanations of natural selection, which we consider to be an adequate definition. Alarmingly, 27.4 percent of students did not mention a single key concept of natural selection in their explanations, and 37.9 percent of students employed only one. Ideas central to natural selection, such as the overproduction of offspring, comprised less than 5 percent of all key concepts mentioned by students. Collectively, these results suggest that prior coursework did not foster a working knowledge of natural selection for most students.

Our results also suggest that even after high school instruction about natural selection and a semester of introductory biology, misconceptions make up a significant part of students' conceptual understanding of natural selection. Students arrived in class employing nearly as many misconceptions in their explanations of evolutionary change as they did key concepts. In addition, while it is deeply worrying that 70 percent of students employed misconceptions in their evolutionary explanations, it is even more alarming that remarkably diverse assemblages of misconceptions ($n > 4$) were present in 40 percent of students (see also Bishop and Anderson 1990). It should be of great concern that only 30 percent of biology majors in the active-learning group (and 14 percent in the postcourse comparison group) lacked misconceptions postcourse.

It appears that misconceptions such as goal-directedness, use and disuse, and acclimation are envisioned by students as more conceptually convincing, understandable, or employable than the key components of natural selection elucidated by Darwin (Mayr 1982). Indeed, when beginning second-semester biology undergraduates are presented with evolutionary problems, they employ misconceptions almost as readily as empirically demonstrated scientific facts and inferences (536 total key concepts versus 403 total misconceptions). These results are concordant with several research studies that demonstrate a substandard understanding of evolution and a wealth of misconceptions in nonmajors and first-semester biology students (Bishop and Anderson 1990, Dagher and BouJaoude 1997).

Pedagogically, it may be necessary to more explicitly explore specific misconceptions and suites of key concepts as alternative explanations for particular evolutionary scenarios, so that students can better appreciate the greater explanatory power of natural selection. Such exercises would be in line with current conceptual-change theories in science education (Strike and Posner 1992). The drawback of conceptual-change approaches is that they often require considerable class time at the expense of other topics. Nevertheless, if we are committed to extinguishing evolutionary misconceptions, such actions may be necessary. Simply put, our study shows that active learning and an evolution-infused introductory biology curriculum provide significant but still woefully limited learning gains.

To address the inadequacy of early evolution education, upper-division coursework must provide additional opportunities for students to enhance their evolutionary knowledge, practice evolutionary problem solving, and devalue misconceptions. Abundant research suggests that if this is being done, it is not successful (Brumby 1984, Affanato 1986, Osif 1997, Nehm and Schonfeld forthcoming). If we are truly committed to instilling an understanding of evolution in college undergraduates, both introductory and advanced coursework will require extensive pedagogical and curricular revisions, along with rigorous assessments to measure the effects of these innovations on students' knowledge and misconceptions of evolution.

Acknowledgments

We thank the working group on evolution at the National Academy of Sciences Summer Institute on Undergraduate Education in Biology for helpful discussions on the issue of teaching evolution in introductory biology. This project was supported in part by a National Science Foundation Early Career Award to R. H. N.

References cited

Affanato FE. 1986. A survey of biology teachers' opinions about the teachings of evolutionary theory and/or the creation model in the United States in public and private schools. PhD dissertation. University of Iowa, Iowa City.

Anderson DL, Fisher KM, Norman GJ. 2002. Development and evaluation of the conceptual inventory of natural science. Journal of Research in Science Teaching 39: 952–978.

Ary D, Jacobs LC, Razavieh A. 2002. Introduction to Research in Education. 6th ed. Belmont (CA): Wadsworth/Thomson Learning.

Bishop B, Anderson C. 1990. Student conceptions of natural selection and its role in evolution. Journal of Research in Science Teaching 27: 415–427.

Bloom B. 1956. Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain. New York: David McKay.

Brumby M. 1984. Misconceptions about the concept of natural selection by medical biology students. Science Education 68: 493–503.

Clough EE, Wood-Robinson C. 1985. How secondary students interpret instances of biological adaptation. Journal of Biological Education 19: 125–130.

Dagher ZR, BouJaoude S. 1997. Scientific views and religious beliefs of college students: The case of biological evolution. Journal of Research in Science Teaching 34: 429–445.

Demastes SS, Settlage J, Good RG. 1995. Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. Journal of Research in Science Teaching 32: 535–550.

Duit R. 2006. Students' and teachers' conceptions and science education. (17 January 2007; *www.ipn.uni-kiel.de/aktuell/stcse/stcse.html*)

Gould SJ. 1999. Rocks of Ages: Science and Religion in the Fullness of Life. New York: Ballantine.

Jensen MS, Finley FN. 1997. Teaching evolution using a historically rich curriculum and paired problem solving instructional strategy. American Biology Teacher 59: 208–212.

Lawson AE, Worsnop WA. 1992. Learning about evolution and rejecting a belief in special creation: Effects of reflective reasoning skill, prior knowledge, prior belief, and religious commitment. Journal of Research in Science Teaching 29: 142–166.

Mayr E. 1982. The Growth of Biological Thought. Cambridge (MA): Harvard University Press.

Nehm RH, Reilly L. 2007. Measuring knowledge of natural selection: A methodological comparison of C.I.N.S., an open-response instrument, and oral interview. Proceedings of the Association of Science Teacher Education 2007 International Conference; 4–7 January 2007, Clearwater Beach, Florida. Forthcoming.

Nehm RH, Schonfeld IS. Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? Journal of Science Teacher Education. Forthcoming.

[NRC] National Research Council. 2000. How People Learn: Brain, Mind, Experience, and School. Exp. ed. Washington (DC): National Academies Press.

———. 2003. BIO 2010: Transforming Undergraduate Education for Future Research Biologists. Washington (DC): National Academies Press.

Osif BA.1997. Evolution and religious beliefs: A survey of Pennsylvania high school teachers. American Biology Teacher 59: 552–556.

Passmore C, Stewart J. 2002. A modeling approach to teaching evolutionary biology in high schools. Journal of Research in Science Teaching 39: 185–204.

Scharmann LC. 1993. Teaching evolution: Designing successful instruction. American Biology Teacher 55: 481–486.

Strike KA, Posner GJ. 1992. A revisionist theory of conceptual change. Pages 147–176 in Duschl R, Hamilton R, eds. Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice. Albany (NY): SUNY Press.

doi:10.1641/B570311
Include this information when citing this material.