Article

Six Classroom Exercises to Teach Natural Selection to Undergraduate Biology Students

Steven T. Kalinowski,* Mary J. Leonard,† Tessa M. Andrews,* and Andrea R. Litt*

*Department of Ecology and †Department of Education, Montana State University, Bozeman, MT 59717

Students in introductory biology courses frequently have misconceptions regarding natural selection. In this paper, we describe six activities that biology instructors can use to teach undergraduate students in introductory biology courses how natural selection causes evolution. These activities begin with a lesson introducing students to natural selection and also include discussions on sexual selection, molecular evolution, evolution of complex traits, and the evolution of behavior. The set of six topics gives students the opportunity to see how natural selection operates in a variety of contexts. Pre- and postinstruction testing showed students' understanding of natural selection increased substantially after completing this series of learning activities. Testing throughout this unit showed steadily increasing student understanding, and surveys indicated students enjoyed the activities.

INTRODUCTION

Evolution is the unifying theory of biology (Dobzhansky, 1973). It also may be the most well-supported scientific theory that is rejected by a large proportion of Americans (e.g., Miller *et al.*, 2006). This high-profile controversy will be familiar to any instructor teaching evolution in the United States. What many biology instructors may not realize is the theory of evolution is also conceptually difficult for students to understand. In particular, natural selection, the main cause of evolution, is a challenging concept for many students. This is ironic, because, as many authors have emphasized (e.g., Coyne, 2009, p. xvi), natural selection is not a complicated process.

The main reason students often have trouble understanding natural selection is that they have misconceptions regarding what causes populations to change. By "misconception," we mean a commonly held idea that is inconsistent with scientific understanding and is resistant to instruction (sensu

Hammer, 1996). The origin and cognitive structures that give rise to student misconceptions are the subject of ongoing research (e.g., Hammer, 1996; diSessa, 2006; Mason, 2007; Abrams and Southerland, 2010), but the actual misconceptions students have are well documented (see Gregory [2009] for an excellent introduction to the large literature on student misconceptions relating to natural selection). Most misconceptions relating to natural selection are variations of the belief that individuals evolve. Students often believe individuals change because they need to, because they want to, because the environment changes them, or because they use or do not use specific body parts—and that these changes are passed on to offspring (e.g., Brumby, 1984; Bishop and Anderson, 1990; Nehm and Schonfeld, 2008). These misconceptions are frequently similar to Lamarck's (1809) pre-Darwinian theory of evolution.

Student misconceptions regarding natural selection are remarkably resistant to instruction (for reviews, see Tanner and Allen, 2005; Sinatra *et al.*, 2008; Gregory, 2009). Bishop and Anderson (1990) showed that half of the students in an introductory biology course left the course with misconceptions regarding natural selection—even though the lectures and laboratory exercises in the course were designed specifically to address misconceptions. Nehm and Reilly (2007) found that 70% of biology majors completing an introductory biology course had at least one misconception regarding natural selection—even though the instructors emphasized evolution as a theme throughout the course and used active-learning exercises extensively to promote learning. Andrews *et al.* (2011b) studied the effectiveness of introductory biology

courses throughout the United States and found that pre- and postinstructional measures of student understanding of natural selection in many courses were statistically identical.

Carefully designed instruction is necessary to help students replace misconceptions with scientifically supported conceptions. There is a consensus among science educators that active learning is more successful than traditional lectures in developing student understanding. Active learning essentially occurs when an instructor stops lecturing and students work on a question or task designed to help them understand a concept (Andrews *et al.*, 2011b). A classic example is a think–pair–share discussion, in which students first individually think about a question posed by the instructor, then pair up with other students to discuss the question, and finally share answers in a whole-class discussion. Students can learn twice as much when lectures contain discussions than when instructors simply lecture (e.g., Hake, 1998; Knight and Wood, 2005; Haak *et al.*, 2011). However, such active-learning instruction may not be sufficient for replacing persistent misconceptions unless instructors design learning activities that specifically focus on helping students move beyond misconceptions to scientifically supported ideas (Duit and Treagust, 2003; Murphy and Mason, 2006; Andrews *et al.*, 2011b). Ongoing research continues to identify the most effective ways to help students recognize and change misconceptions (e.g., Vosniadou, 2008). Nonetheless, there is a consensus that the following instructional practices are useful. First, instructors must help students become aware of what they believe and how to recognize which of their ideas conflict with observations or with previous knowledge (Hewson *et al.*, 1998; Bransford *et al.*, 2000). Second, students are more likely to learn any concept, including ones for which they have misconceptions, when they study multiple examples of the concept at work (Catrambone and Holyoak, 1989; Mestre, 2003; Marton, 2006). Third, helping students construct scientific conceptions involves introducing them to scientists' "ways of seeing" (Scott *et al.*, 1991); in this case, seeing situations of natural selection in terms of its genetic basis (Kalinowski *et al.*, 2010).

There is a shortage of classroom exercises for teaching natural selection that require the active participation of students, that have been designed specifically to deal with misconceptions, and that have been shown to be effective in promoting a deep understanding of natural selection. In this investigation, we developed and assessed six classroom exercises for teaching natural selection to undergraduate students. The primary learning goal for the exercises was for students to thoroughly understand natural selection.

METHODS

Participants and Context

We designed and assessed (with institutional review board permission) six lessons relating to natural selection in an introductory course on ecology and evolution during 2011 and 2012. This course was the third in a three-semester sequence of introductory biology courses for biology majors at Montana State University. The previous courses in this sequence covered physiology, cell biology, biochemistry, and genetics. Classes met three times a week for a 50-min lecture and once a week for a 3-h lab. Twenty lectures in the course were devoted to studying evolution, including six focused on natural selection.

The student body changed during our study. In 2011, 41 students enrolled in the course: 85% were freshman, 12% sophomores, and 3% seniors. In 2012, 47 students were enrolled in the course: 54% were sophomores, 29% were juniors, and 17% were seniors. The large proportion of freshman in 2011 was due to an advising error, but provided us with an unexpected opportunity to test our lessons on two different sets of students. In both years, virtually all students reported by raised hands they were preparing for careers in the health sciences.

We did not collect demographic data on race either year. However, the student body in the College of Letters of Sciences at Montana State University is 90% Caucasian, and there is no reason to suspect enrollment in this course was substantially different.

Two instructors taught the six learning activities described here. In 2011, S.T.K. taught the entire course. S.T.K. is a tenured professor, who had eight years of teaching experience in 2011. He had taught this course six times previously. In 2012, T.M.A. taught the six lectures that included the learning activities described in this paper. (S.T.K. taught the remainder of the course). At the time she taught these classes, T.M.A. was a fourth-year PhD student whose primary teaching experience consisted of serving as a teaching assistant for this course the previous year. Half of T.M.A.'s PhD research related to undergraduate biology education (Andrews *et al.*, 2011a, 2011b), so she likely had more pedagogical knowledge than most graduate students.

Description of the Six Learning Exercises

We created six learning exercises (Table 1) to teach students how natural selection works and to help correct common misconceptions regarding its operation. Our exercises focused on students' ideas, drawing from principles of active learning and conceptual change teaching. Although exercises varied, they generally included the following elements: we began by asking a thought-provoking question that asked students to account for a situation in which natural selection was occurring, to evaluate whether it was occurring, or to predict what would happen in the situation. We elicited students' initial ideas in response to the question, either in writing or as a class discussion. We provided opportunities for students to discuss their ideas with their peers. We explicitly discussed misconceptions in ways intended to help students evaluate them; for example, describing circumstances for which a misconception could not provide an adequate explanation, facilitating students in identifying pros and cons of an explanation, and having students compare alternate explanations. Finally, we emphasized the genetic basis of natural selection by providing genetic data. The learning exercises, described below, were given in six lectures (one per lecture) in the order described below. Two of these exercises have been described previously: Kalinowski *et al.* (2010) described the dog-breeding discussion, and Andrews *et al.* (2011a) described and assessed the human evolution discussion.

Introduction to Selection: Dog Breeding. In the first chapter of *On the Origin of Species*, Darwin described how plants and animals have been bred to have desired characteristics

Table 1. Focal questions for six classroom exercises for teaching natural selection, and the approximate amount of time required to discuss each question

Exercise	Focal question	Time required
Dog breeding	How could you turn a pack of wolves into Chihuahuas? Explain why your strategy would work.	30 min
Coat color in Oldfield mice	Where did the gene for white fur come from?	10 min
Human evolution	Are humans evolving? If so, which trait is changing? Explain.	50 min
Peacock trains	How would you test whether peacocks with large spots on their trains have "good genes"?	15 min
E. coli antibiotic resistance	What does natural selection predict about the evolution of antibiotic resistance in *E. coli*?	5 min
Lemming suicide	Could suicide be an adaptation in lemmings?	15 min

and used such "artificial" selection as an analogy to introduce how "natural" selection works in nature. The analogy is still powerful, and we used it to introduce our students to natural selection. We began this discussion by telling the class that all dog breeds are descended from wolves and asking the class "If you had a bunch of wolves and wanted a Chihuahua, how would you create one?" Students discussed the question in pairs and we then elicited answers from randomly selected students. A "correct" answer for this question is that Chihuahuas can be bred from wolves by selectively breeding small wolves with short faces and wiry tan hair for many generations. Some students provided this answer, but many proposed raising wolves in a warm environment "so they will not need such heavy fur" and providing them with plenty of food "so the wolves become less aggressive and develop smaller teeth." Such responses reveal the misconceptions that the environment causes individuals to evolve and traits evolve from their use or disuse.

Once we recorded a diversity of answers on the board, and the class could see the need to reconcile the differences expressed, we asked students to comment on the feasibility of each proposal. This created some confusion. Some students clearly understood why raising wolves in domestic environments will not cause them to become Chihuahua-like. Others did not. We resolved the confusion by making a connection to genetics and reminding the class that wolf pups grow up to be adult wolves because they have wolf genes, not because they are raised in a forest hunting elk.

We emphasized that the selective-breeding program will work, because when a breeder preferentially chooses small, tan wolves to breed, he or she is selecting wolves with specific DNA sequences. We showed the class DNA sequences for one of the genes that differentiate Chihuahuas from wolves. Body size in wolves and dogs is influenced by variation at the insulin-like growth factor 1 (*IGF1*) gene. Chihuahuas (and other miniature breeds) have an adenine at position 44,228,468 in the dog genome, and wolves have a guanine (Sutter *et al.*, 2007).

After we completed the dog-breeding discussion the instructor gave a short lecture describing the requirements for natural selection. We emphasized evolution by natural selection would occur if: 1) there was phenotypic *variation* in a population; 2) this variation was *heritable*, and 3) this variation influenced the *reproductive success* of individuals. We repeatedly emphasized these three concepts in all our subsequent lectures, and later used them in our assessment of student learning.

Source of Variation: Coat Color in Oldfield Mice. One of the challenges for students studying natural selection is that the genetic basis of evolution is invisible. Therefore, discussing how natural selection affects the frequency of DNA sequences in a population should be useful. Coat color in Oldfield mice, *Peromyscus polionotus*, makes a useful case study for how natural selection might work in the wild.

We began our discussion of coat color in Oldfield mice by describing the natural history of the species. Oldfield mice live in the southeastern United States and generally have dark fur that provides camouflage from owl predation for mice living in forests. In contrast, a subspecies of Oldfield mice, *P. p. leucocephalus*, lives on the white sand dunes of Santa Rosa Island on the Gulf Coast of Florida and has nearly white fur. Experiments have shown that this coloration protects mice from owls (Kaufman, 1974). The difference in fur color is largely caused by a single nucleotide change in the melanocortin-1 receptor gene. Mice with a thymine at a specific location in the gene have much lighter fur than mice with a cytosine at that location (Hoekstra *et al.*, 2006). The mice on Santa Rosa Island are the only mice in the southeastern United States with this genetic variant.

After we presented the preceding material to the class, we asked our students to consider a plausible history for Oldfield mice on Santa Rosa Island. Specifically, we asked them to assume that Oldfield mice colonized Santa Rosa Island when sea levels were lower during the last ice age, and that all of the mice colonizing the island had brown fur. We emphasized that being a brown mouse living on white sand dunes put the mouse in danger. We then asked our students "How did the population of brown mice become white?" Students thought about this question and wrote an answer on an index card. Then students discussed their answers in small groups, after which we solicited answers from the entire class. After collecting answers from several groups, we discussed the importance of mutation in creating new phenotypes, and emphasized that mutations randomly changed DNA sequences—and do not necessarily make the changes that organisms want or need.

Natural Selection: Human Evolution. As we have discussed above, many students have persistent misconceptions regarding how natural selection operates. Our human evolution discussion is particularly effective at eliciting such misconceptions. Andrews *et al.* (2011b) described in detail how we conduct this discussion, and we will refer the reader to their

paper for a full description of the exercise. A brief description of the discussion should be adequate here.

We begin the discussion by asking our students "Are humans still evolving? If so, what trait is changing? Explain why or why not." Students discussed this question in groups and then presented answers to the class. Many of the answers lacked any reference to the principles of natural selection. For example, students answered "people are getting balder" or "computers are making us smarter." We then discussed the likelihood of such changes occurring, given the requirements for natural selection (variation, heritability, differential reproductive success).

Sexual Selection: Peacock Trains. It should not be too difficult for students to understand how natural selection can change the coat color of populations of mice living on sand dunes, but the origin of other traits is harder to understand. For example, Darwin struggled to figure out how natural selection could have created highly ornamental traits, such as the peacock's train, that would seem to reduce an individual's ability to survive. In 1860, he wrote Asa Gray: "the sight of a feather in a peacock's tail . . . makes me sick!" Darwin eventually proposed elaborate peacock trains could evolve if peahens preferred mating with peacocks that had elaborate trains. Students are likely to experience the same confusion as Darwin, so discussing peacock trains provides an ideal opportunity to reinforce how natural selection works, and to introduce the topic of sexual selection

We began a discussion of sexual selection by pointing out that natural selection is often summarized as "survival of the fittest" but noted that many animals have traits that seem to decrease their chances of survival. We proposed that the elaborate trains of peacocks are perfect example of such traits; the long feathers would seem to increase the risk of predation. We then asked our students to come up with as many hypotheses as possible to explain why peacocks have such elaborate trains. Students proposed that tail feathers scared away potential predators or helped attract mates. We are not aware of any research on the ability of peacock trains to deter predators, so we told our students that we were going to restrict our discussion to the mating preference hypothesis.

Next, we showed our class a graph from Petrie's famous mate choice study at the Whipsnade Zoo (Figure 3 in Petrie *et al.*, 1991). This graph shows that peacocks with many eyespots (a sign of an elaborate train) garnered more mates than males with fewer eyespots. We also showed students that peahens mated more often with peacocks with bright tail feathers (Figure 3 in Loyau *et al.*, 2007) and mated less often with males that had eye spots removed by researchers (Figure 3 in Petrie and Halliday, 1994). These data answer the question of why peacocks have elaborate trains: elaborate trains increase their chances of mating. We explained how such a mating preference could give rise to the elaborate trains using the requirements for natural selection that we used throughout the course (i.e., natural selection requires variation, heritability, and differential reproductive success).

The previous discussion gave rise to a new, interesting question, namely "Why do peahens prefer to mate with peacocks with elaborate trains?" We let our students discuss this for a couple of minutes, solicited answers, discussed possible explanations, and concluded that peacocks with elaborate trains have much better genes than peacocks with less

elaborate trains. We introduced this to our students as the "good genes hypothesis" and asked them how they would test it. This was a difficult question for our students. Many students suggested examining peacocks with elaborate trains and testing them to determine whether they were healthier or otherwise superior to peacocks with less elaborate trains. The problem with this experimental design is that both traits (elaborate trains and health) are likely influenced by environmental factors. Peacocks raised in favorable conditions might have elaborate trains and be healthy—and not have good genes. A rigorous test of the good genes hypothesis must test whether peacocks with elaborate trains contribute better genes to offspring than peacocks with less elaborate trains. Petrie (1994) did this by randomly mating peahens and peacocks and showing that the chicks of peacocks with elaborate trains were more likely to survive than the chicks of peacocks with less elaborate trains (see Figure 2 in Petrie, 1994). After we completed this discussion, we formally defined sexual selection, talked about why females get to choose, and discussed the consequences of sexual selection.

Evolution of a Complex Trait: Antibiotic Resistance in **Escherichia coli.** Students frequently have a hard time understanding how complex traits, such as the vertebrate eye, have evolved through a combination of random mutation and natural selection. Darwin (1859) anticipated this. He wrote in Chapter 6 of the *Origin of Species*: "To suppose that the eye . . . could have been formed by natural selection, seems, I freely confess, absurd in the highest possible degree." Darwin then went on to explain how such complex structures could evolve via natural selection through the accumulation of small changes, so long as each small change improved the ability of the organism to survive and reproduce. This is an important concept for students to understand.

We began our discussion of the evolution of complex traits by introducing our students to antibiotic resistance in *E. coli*. *E. coli* are rod-shaped bacteria that live in the human digestive tract. Most strains are harmless, but some cause infections that doctors may treat with penicillin or other antibiotics. Some *E. coli* have a form of the enzyme β-lactamase that can break down the β-lactam ring present in penicillin and other antibiotics, including cefotaxime (Baquero and Blazquez, 1997). The degree to which *E. coli* are resistant to antibiotics depends on how quickly β-lactamase can break down the β-lactam ring of the antibiotic, and this rate depends on the sequence of amino acids in the β-lactamase enzyme. The β-lactamase allele most commonly present in *E. coli* populations is the TEM-1 allele, which provides a modest amount of resistance to the antibiotic cefotaxime. In contrast, the TEM-52 allele provides more than 4000 times as much resistance to cefotaxime. TEM-52 differs from TEM-1 by three amino acid substitutions.

Once we introduced our students to this case study, we asked them "What does the mechanism of natural selection predict about the evolution of TEM-1 β-lactamase to TEM-52 β-lactamase?" We let our students discuss this question in pairs, and then randomly called on students for answers. The answer to the question is that TEM-1 must evolve into TEM-52 via three mutations and each mutation must increase the ability of *E. coli* to survive and reproduce. Our students have not had a difficult time deducing this. Weinreich *et al.* (2006) synthesized *E. coli* with each possible combination of amino acids

at the three sites that needed to be changed and showed that there were indeed sequences of mutations that could change TEM-1 to TEM-52, while increasing antibiotic resistance in each step. We completed this discussion by telling students that this property of β-lactamase evolution—that each mutation must increase fitness—is a general feature of evolution, and the lecture continued with a discussion of the evolution of the vertebrate eye.

Evolution of Behavior: Apparent Suicide in Lemmings. A common misconception among students is that natural selection favors traits that are "good for the species." The lemming suicide myth provides an ideal opportunity to discuss why this is not always true and to show students that the evolutionary origins of altruistic behaviors require special explanation. We began our discussion of lemming suicide by telling students we were going to show them a nature documentary about lemmings, and after the film, they would be asked to answer the question: "Could suicide be an adaptation in lemmings?" Then we showed the class a 3.5-min film clip from the 1958 Walt Disney documentary *White Wilderness*. The film clip is available on YouTube (www.youtube.com/watch?v=xMZlr5Gf9yY or search www.Youtube.com for "White Wilderness"). The film appears to show lemmings jumping off cliffs into the Arctic Ocean and swimming to their death. The narrator does not explicitly claim the lemmings are committing suicide, but strongly implies this is what is happening, and most viewers will believe this is what they are watching.

After watching the film, we asked our students "Could suicide be an adaptation in lemmings to prevent overcrowding?" and let them discuss this in small groups. Then we randomly called on students and discussed the answers we obtained. Through discussion, we attempted to explain that natural selection is unlikely to favor suicide, because individuals that commit suicide will not pass on their genes. Gary Larson's *Far Side* cartoon showing a bunch of lemmings rushing into the water, including one wearing a life preserver, was useful for showing how selfish lemmings who did not commit suicide would pass on their genes and cause the frequency of selfish behavior to increase.

Our students seem to quickly understand why self-destructive behavior is unlikely to evolve via natural selection, but struggled to explain the apparent mass suicide they witnessed in the documentary. For example, some students proposed the lemmings were committing suicide after they had reproduced or because they had a disease they did not want to transmit to their offspring. Others suggested that the lemmings were leaving overcrowded locations in search of a better place to live and reproduce and happened to come across a body of water in their way. These are reasonable suggestions, but turn out to be incorrect. The actual explanation reveals more about human nature than natural selection. According to a documentary aired by the Canadian Broadcasting Company on May 5, 1982 (www.cbc.ca/fifth/cruelcamera/video2.html), the dramatic footage in the film *White Wilderness* of lemmings jumping into the Arctic Ocean was faked. Apparently, the lemmings were pushed off of a cliff into a river. And yet, the film won the 1958 Academy Award for best documentary.

Assessment

We used two instruments to measure how well students understood natural selection before, during, and after our series of six exercises. The first instrument was a 10-question version of the Conceptual Inventory of Natural Selection (CINS-abbr; Anderson *et al.*, 2002; Fisher, Williams, Lineback, and Anderson, personal communication; see Table 3 below). This is a multiple-choice test with distracters designed to appeal to students having common misconceptions regarding natural selection and related concepts. In addition, we used seven short essay questions (Table 2) to give students an opportunity to answer evolutionary questions in their own words. The questions we used (Table 2) were variations of questions developed by Bishop and Anderson (1990) for their Open Response Instrument and are similar in form to questions recently described by Nehm *et al.* (2012). Each question asked students to explain how an adaptation in a familiar animal might have evolved. Six of the seven questions involved the gain of a trait, and one involved the loss of a trait.

Students were required to complete the CINS-abbr and the short essay questions, but the accuracy of their responses did not affect their grades. Students were told, however, that their responses would be evaluated as part of their course participation grade—which comprised 2% of their grade for the course.

We used the CINS-abbr to measure how much students learned during our entire unit on natural selection. We administered the test twice: in the class period before we began our unit on natural selection and in the class period after we completed the last of our six learning exercises. We calculated the average score on the CINS-abbr before and after instruction and used this to calculate the normalized gain (Hake, 1998) for the class. We did this in 2011 and 2012. We tested the statistical significance of increases in test scores using a nonparametric sign test. Scores from students who did not take both the pre- and postinstruction tests were not included in the analysis.

We used seven short essay questions (Table 2) to assess learning in two ways. First, we used the essay questions as an alternative to the CINS-abbr to measure how much students learned in the entire unit on natural selection. Second, we used these essay questions to monitor learning throughout our six-lecture unit on natural selection.

We administered the essay questions before, during, and after our unit on natural selection as follows. Each student answered one short essay question before we began our unit on natural selection and then one question after each of the six classroom exercises. We were concerned the questions might vary in difficulty, so we randomly divided our class into seven groups and gave each group a different question each day we tested the class. Group one answered question one on the first day (before instruction), question two on the second day, and so on, answering question seven after the last of the learning exercises. Group two answered question two on the first day, question three on the second day, and question one after instruction was over, and so on. With this design, each student answered a different question each day, but the class as a whole always answered the same set of questions. This allowed us to compare the average score of the class on different days. Administering these short essay questions took a substantial amount of class time, so we did this only in 2011.

Table 2. Short essay questions used to assess student understanding of natural selection

Cheetahs are able to run faster than 60 miles per hour when chasing prey. How would a biologist explain how the ability to run this fast evolved in cheetahs, assuming their ancestors could only run 20 miles per hour?

Polar bears have white fur that blends in well with their snowy surroundings. This helps polar bears stalk and hunt seals. Polar bears are believed to have evolved from bears that had brown fur. How would a biologist explain how the white fur of polar bears evolved from bears with brown fur?

Musk oxen are large animals that that live in the coldest parts of the Arctic and look something like shaggy cows. Musk oxen have the warmest wool of any mammal. How would a biologist explain how musk oxen evolved this warm wool, assuming that their ancestors had wool that was less warm?

Flying squirrels have folds of skin between their front and back legs that allow them to glide (although not fly) between trees. How would a biologist explain how flying squirrels evolved these folds of skin, assuming their ancestors did not have these folds?

Eagles have keen eyesight that allows them to spot mice and other prey while soaring high above the ground. How would a biologist explain how eagles evolved their keen eye sight, assuming their ancestors had less keen eyesight?

Camels store fat in their humps, which allows them to travel for long distances without eating. How would a biologist explain how camels evolved their humps, assuming their ancestors did not have humps?

Whales are large mammals with streamlined bodies that allow them to swim easily in the ocean. Unlike most mammals, whales do not have hind limbs. How would a biologist explain how whales lost their hind limbs, assuming their ancestors had hind limbs?

We graded all of the short essay questions using the rubric of Andrews *et al.* (2011b). This rubric evaluated student responses according to how well they addressed the three key concepts we emphasized in lecture: variation for a trait within a population, heritability of the trait, and differential reproductive success. We did not discuss the questions or student answers to the questions during class, and we did not read any of the student answers until after the course was over. We scored student responses in a random order with the name of the student and the date obscured. All of the responses were independently scored by two of the authors (S.T.K. and T.M.A.), and any differences in scores were resolved through discussion.

We used the short essay data and an analysis of variance (ANOVA) to estimate how much students learned during the first lesson. To do this, we compared the average test score

Table 3. Frequency of the most commonly selected wrong answer in the CINS-abbr before (PRE) and after instruction (POST) instruction in 2011

	Frequency	
Question and most popular wrong answer before instruction	PRE	POST
1. Once a population of finches has lived on a particular island for many years in a relatively steady climate a. The population size continues to grow rapidly, at maximum rates.	0.21	0.06
2. What is the best way to characterize the evolutionary changes that occur in a finch population? d. The environment causes specific mutations in individual finches to help them survive and reproduce.	0.32	0.08
3. Imagine that the ancestors of the cactus finch colonized a new island. They did not have the beak type seen in the cactus finch today. How did the cactus finches' unique beak type first arise? c. In the descendants of the original birds, the environment gradually caused the genetic changes that were necessary to live on that island.	0.15	0.08
4. What type of variation in a finch population is passed to the next generation? d. Any characteristics that were positively influenced by the environment.	0.35	0.08
5. What caused populations of finches having different beak shapes and sizes to become distinct species? d. The environment of each island gradually molded beak shape in the new species, because that particular shape was needed to obtain the food.	0.50	0.00***
6. A typical natural population consists of hundreds of guppies Which statement best describes the guppies of a single species in an isolated population? b. The guppies share all of the essential characteristics of the species; the minor variations they possess do not affect survival or reproduction.	0.41	0.22
7. Once a population of guppies has been established for a number of years in a pond with other organisms, including predators, what will likely happen to the population size, assuming that conditions remain relatively constant. c. The guppy population size will gradually decrease.	0.12	0.06
8. What is the best way to characterize the evolutionary changes that occur in a guppy population over time? d. Mutations occur to meet the changing needs of the guppies, because the environment changes.	0.35	0.00***
9. Where did variation in spot colors and patterns in guppy populations come from? a. The guppies needed particular spots to survive and reproduce, so those spots developed.	0.24	0.08
10. What could cause one guppy species to change into three species over time? c. Different mutations in each environment occurred, because they were needed, and each population gradually became a new species.	0.47	0.06***

*p-value (calculated from Fisher's exact test) < 0.05.

**p-value < 0.01.

***p-value < 0.001.

Table 4. Average responses of surveys used to measure student attitudes toward six learning exercises[a]

	This activity . . .			
	held my interest	challenged me intellectually	was under-standable to me	was a valuable learning experience
Dogs	4.9	4.5	5.5	5.1
Mice	4.6	4.1	5.4	4.7
Humans	4.8	4.6	5.4	5.0
E. coli	4.8	4.6	5.6	5.0
Peacocks	—	—	—	—
Lemmings	5.4	5.5	5.3	5.2

[a]Student responses were coded on a scale from 1 to 6: 1 = strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = slightly agree, 5 = agree, 6 = strongly agree. Surveys were conducted on the same day each learning exercise was performed. Survey data for the Peacock exercise were not collected, due to a lack to time on the day the exercise was taught.

collected before instruction (PRE) with the average score after the first lesson (dog breeding), using ANOVA. We explicitly paired scores for each individual student and accounted for potential differences in difficulty among the various questions administered. We also evaluated an interaction term (i.e., = difference between PRE and the first lesson × essay question) to determine whether the degree of initial learning gains differed based on the essay question assigned; we removed the interaction if it was not statistically significant.

Next, we were interested in determining whether and the degree to which student scores on the essay questions increased throughout our unit. In this second analysis, we analyzed only the data collected after the first lesson (dog breeding) and used a linear regression to quantify the degree of learning gain. We did not include the test scores collected before instruction (PRE), because we wanted to see whether students benefited from continued instruction (i.e., represented by a nonzero and increasing slope of the line) or whether the gains after initial instruction (which should be highest) were minimal (i.e., represented by a flat line). We analyzed these data with a generalized linear mixed model that allowed us to account for repeated observations from the same students by including a compound symmetric covariance structure (Littell *et al.*, 2006). We also accounted for potential differences in difficulty among the various questions administered. Additionally, we evaluated an interaction term (i.e., = lesson × essay question) to determine whether changes in the average score over the series of exercises in the unit differed based on the essay question assigned; we removed the interaction if it was not statistically significant.

We also used surveys to assess student attitudes toward the six learning activities (see Table 4 below). Participation in these surveys was voluntary and anonymous.

RESULTS

Data from the CINS-abbr showed that, before instruction, many students appeared to have misconceptions regarding natural selection (Table 3). Between 25 and 50% of our class seemed to believe evolution was caused by the environment changing individuals or species evolving out of need (Table 3). As we discussed in the *Introduction*, these are common misconceptions among introductory biology students.

Our class showed impressive learning gains on the 10-question version of the CINS both years we administered this test (Figure 1). In 2011, pre- and postinstruction CINS-abbr scores were available for 32 of the 41 students in the course. Before instruction, the class average on the CINS-abbr was 6 (out of 10; $\hat{\sigma} = 2.55$). After instruction, the average was 8.88 ($\hat{\sigma} = 1.45$). This corresponds to a normalized gain of 0.72. Analysis of student responses on specific questions suggested the frequency of specific misconceptions declined after instruction (Table 3). In particular, fewer students seemed to have the misconception that evolution happens because the environment changes individuals.

We observed similar learning gains on the CINS-abbr in 2012. In 2012, pre- and postinstruction CINS-abbr scores were available for 42 of the 47 students in the courses. The average score among these students increased from 7.14 ($\hat{\sigma} = 2.55$) to 9.24 ($\hat{\sigma} = 1.03$) which corresponds to a normalized gain of 0.73. As we discuss below, these are exceptionally large learning gains for this test. Notice also that the normalized gain was essentially the same both years (0.72 vs. 0.73), despite the fact the natural selection unit was taught by two different instructors and that the composition of the student body was substantially different in 2011 and 2012.

We administered the short essay questions seven times to our class of 41 students in 2011 and received 236 student responses. This corresponds to a response rate of 236/(41 × 7) = 82%. As indicated above, all questions were graded by two researchers; the correlation between scores was initially 0.82 (all differences resolved via discussion) Seven student responses were culled before we began statistical analysis. Six students missed three lectures in a row, and we dropped the responses of these students after these absences. This removed six responses from our data. We also dropped one student's response to the final question. It was the only response on the last two testing dates that received a score of zero. The response was unusually short, and the student had earned full credit on two previous questions. Furthermore, this response was identified as an outlier in a general linear model. After dropping these seven responses, we had a final data set of 229 responses, or a response rate of 80%.

Student responses on the short essay questions showed substantial learning gains. Before instruction, the most common score in the class was 0 points (out of 6) and the average score in the class was 2.84. After instruction, the most common score in the class was 6 points, and the average score was 4.90. This corresponds to a normalized gain of 0.65, which, as we discuss below, compares favorably with results from other courses for similar questions.

On the basis of our analysis of student responses to the short essay questions, we quantified a large increase in understanding after the first learning exercise (i.e., difference between PRE and the Dog lesson, Figure 2a). Average scores increased by 1.23 points (95% CI = 0.48–1.98) after the first learning exercise ($F(1, 30) = 11.18$, $p = 0.002$; difference between PRE and Dog lesson in Figure 2).

We also observed that student scores continued to improve throughout the series of our exercises by an average of 0.15 points (95% CI = 0.05–0.25) with each additional exercise
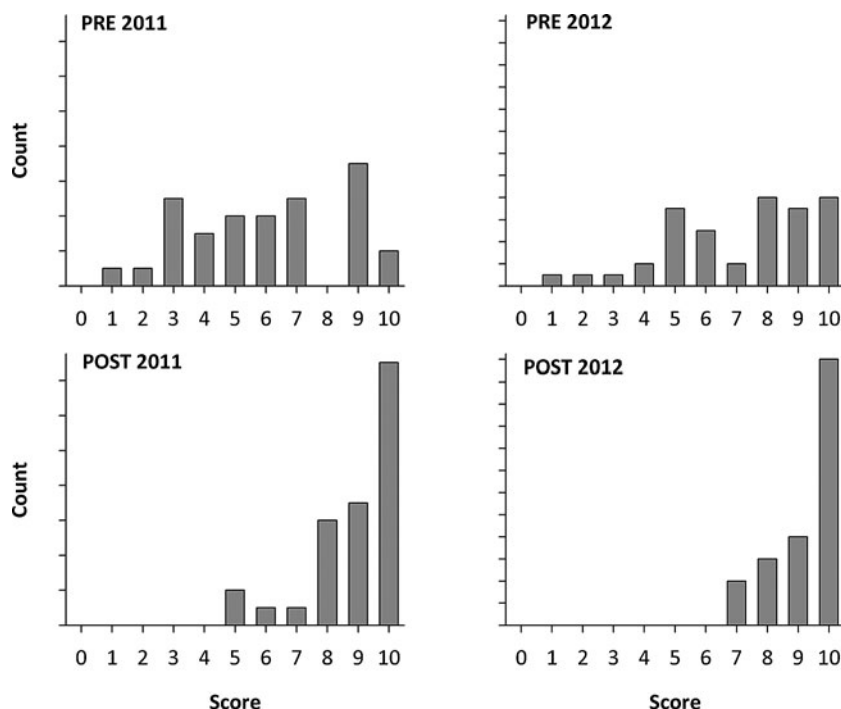
Figure 1. Pre- and postinstruction test scores on the CINS-abbr in 2011 and 2012.

($F(1, 145) = 9.02$, $p = 0.003$; the Dog lesson through the Lemmings lesson on Figure 2a), which amounts to an average total increase of 0.75 points (95% CI = 0.25–1.25) over the series of five exercises. We found little evidence that the direction or magnitude of these increases differed based on which essay question students were assigned (i.e., interactive effect = lesson × essay question, initial increase: $F(6, 24) = 0.61$, $p = 0.72$; subsequent increase: $F(6, 139) = 1.1$, $p = 0.37$).

Our rubric for the short essay questions contained three components: variation, heritability, and differential reproductive success. Student answers consistently received the fewest points for the heritability component (Figure 2b). This component also seemed to be the most difficult for students to learn.

Students' scores for variation and differential reproductive success quickly rose after the first lesson, and then changed relatively little (especially for the concept of variation). In contrast, scores for the heritability component of student answers steadily increased throughout instruction. This increase was responsible for 77% of the increase in scores after lesson 1.

The survey of student attitudes regarding the six natural selection exercises showed that students generally viewed these lessons positively (Table 4). On average, the class agreed the exercises were interesting, challenging, understandable, and valuable learning experiences. None of the exercises received remarkably low or high evaluations, but the lemming discussion (which was the last one taught) did seem
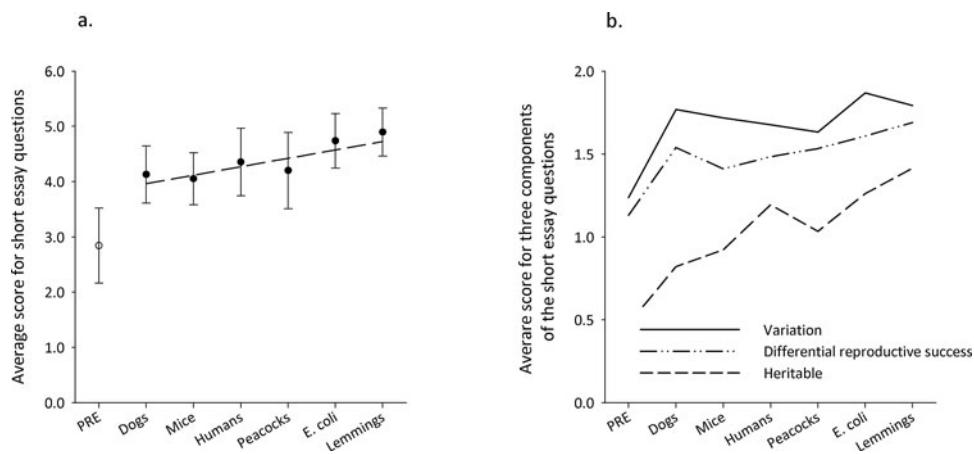

Figure 2. Average scores on the short essay questions before instruction (PRE) and after each lesson (Dogs, Mice, etc.) in 2011. (a) The average score in the class; (b) the average score for each component of the total score included in the rubric. We analyzed the data in (a) in two ways. First, we quantified the initial learning gains after one lesson (difference between PRE and Dogs). Next, we determined whether learning gains continued to occur after initial instruction; the line in (a) shows the predicted increase in the average score for the class throughout the set of exercises. Test scores before instruction (PRE) were not included in this second analysis.

DISCUSSION

We developed and assessed six classroom activities for teaching natural selection to introductory biology students. All our assessments suggested the activities were both engaging and effective.

The learning gains we observed on the 10-question version of the CINS compare very favorably with results from other classrooms. Andrews *et al.* (2011b) administered the same 10-question CINS-abbr to 33 introductory biology courses randomly sampled from major universities across the United States. Students in these classes took the CINS-abbr before and after instruction on natural selection. The average normalized gain among these courses was 0.26, and the highest normalized gain observed was 0.68 (T.M.A., unpublished data). In our classroom, we observed a normalized gain of 0.72 when S.T.K. taught the course and 0.73 when PhD student taught the lessons. Notice that our learning gains were higher than in any of the other classrooms studied by Andrews *et al.*, even when our unit on natural selection was taught by a graduate student with very little teaching experience. We interpret this as evidence that the lessons we describe here are effective for promoting learning.

Results from assessment using short essay questions also compare favorably with results from other institutions. The normalized gain associated with our six learning activities, as measured by our short essay questions, was 0.65. The average normalized gain for a short essay question in the national study of Andrews *et al.* (2011b) was 0.06, approximately one-tenth of what we observed in our classroom. This is additional evidence that our learning exercises were effective. There is, however, a methodological difference to note between our assessment and that in the survey of Andrews *et al.* (2011b). We used seven short essay questions. Andrews *et al.* (2011b) used only one question: a question on cheetahs very similar to our cheetah question (Table 2). The data, therefore, are not perfectly comparable. However, there is no reason to suspect our set of questions was easier than the cheetah question used by Andrews *et al.* (2011b). We observed a standardized gain of 0.88 for the cheetah question in our classroom; the standardized gains for all of the other questions were *lower*. Therefore, the set of questions we used to assess our learning exercises was probably harder than the question used by Andrews *et al.* (2011b).

Recent research (Nehm and Ha, 2010; Nehm *et al.*, 2012) has shown the ability of students to answer questions regarding natural selection depends on the context of the question. In particular, Nehm and Ha (2010) showed that questions that involve trait loss, unfamiliar species, or evolution between species are more difficult for students to answer. Had this research been available when we started our project, we would have broadened the scope of the short-answer questions in our assessment. However, we did include one question relating to trait loss (in whales), and this provides us with some insight for how effective our learning exercises might be for allowing students to reason about natural selection in contexts beyond evolution of a new trait within a species. As expected from Nehm and Ha's (2010) results, scores on our trait-loss question were, on average, lower than the six trait-gain questions. However, the learning gains observed with the trait-loss whale questions did not appear to be lower than those of the other questions. We calculated the normalized gain for each of the seven questions, and the normalized gain for the whale evolution question was 0.60. This was equal to the median of the normalized gain among our seven questions. We never discussed or mentioned trait loss at any time during the course, so it seems likely that our students learned natural selection well enough that they were able to reason effectively about the relatively novel evolutionary scenario of how traits are lost.

Our pre- and postinstruction testing has shown that our students made impressive progress understanding one of the most difficult concepts in biology. We attribute these learning gains to our classroom exercises and suggest four characteristics of our lessons *may* have been important for promoting learning. First, we designed these lessons to specifically target student misconceptions. We did not just use active learning—we used specific methods in an active-learning format to help students move past misconceptions. Second, we explicitly discussed the genetic basis of evolution. Many student misconceptions regarding natural selection relate to genetics, so making a connection to students' prior genetics knowledge should be helpful. Third, we presented students with multiple examples of selection at work in a variety of different contexts. Finally, we did our best to create exercises that engaged our students as deeply as possible (Table 4).

We have attributed the dramatic increase in test scores we observed with our students to the classroom activities we conducted, and have proposed four reasons why our lessons may have been effective. We do acknowledge, however, the design of our study makes it impossible to unambiguously attribute student learning to any specific element of our instruction. We conducted six classroom activities during our unit on natural selection, but these activities were not the only opportunities students had to learn. Students were assigned textbook readings before each lesson, and this may have contributed to some of their learning. Similarly, most of our lectures included a fair amount of ordinary lecturing, and this could have contributed to student learning. There are many other possibilities. However, it is not credible to us that any of these aspects of our course were unique enough to explain the uniquely high learning gains we observed.

Readers may wonder whether any of our six lessons were more or less effective than others. For example, the average class score on the short essay questions dropped after the mice and peacock discussions, which raises the question of whether these lessons were effective or not. Unfortunately, our study was not designed to measure how much students learned in a specific activity. There are several reasons why not. First, we did not perform pre- and postinstruction testing for any of the individual learning activities. We administered the CINS-abbr before and after our unit on natural selection and administered essay questions to the class before, after, and during our unit on natural selection, but we did not perform any assessment immediately before and/or immediately after any of the activities described here. Second, our course had only 41 students, and they answered a single essay question each day we did testing during the unit. This does not give us a lot of statistical power to measure how much students learned from one lesson to the next. Third,

student understanding can temporarily "dip" after a lesson, especially if a lesson presents students with a principle in a new context. When instructors do this, students need to restructure their knowledge to deal with the new context (e.g., Bransford and Schwartz, 1999; Barnett and Ceci, 2002; Lobato, 2008). Fourth, it is important to remember that the essay questions provide only one measure of how well students understand natural selection. Learning gains might have looked different had we used other questions. Finally, we emphasize that our data describes learning gains for the sequence of exercises as we taught them. The apparent effectiveness of any lesson almost certainly strongly depends on what is taught before it. For example, the dog-breeding lesson was associated with higher learning gains than the discussion of beach mice. If we used the mice lesson first and followed it up with a discussion of dog breeding, we might observe the opposite.

Instructors considering incorporating our learning exercises into their courses may wish to modify them to suit the specific needs of their courses. If this is done, instructors must ensure these activities give students the opportunity to personally construct an understanding of natural selection. In particular, we strongly encourage instructors to make sure that students: 1) formulate an initial answer to each question on their own, 2) discuss their answers with a few other students, 3) participate in a discussion with the entire class, and 4) recognize how these activities relate to the main concepts taught in a course. All of this takes time but is important for learning. There is accumulating evidence that instructors attempting to use active-learning methods do not include these elements in their instruction, and therefore fail to see the learning gains they hope to achieve (Turpen and Finkelstein, 2009; Andrews *et al.*, 2011b; Ebert-May *et al.*, 2011).

Instructors using the exercises described in this paper in their classroom should make an effort to assess whether their students are learning natural selection. Some sort of pre- and postinstruction testing is necessary to do this. The CINS-abbr (Anderson *et al.*, 2002), ACORNS (Assessing Contextual Reasoning about Natural Selection) (Nehm *et al.*, 2012), or open-response questions (described by Bishop and Anderson, 1990) are reasonable instruments to use. There are no established criteria for what constitutes acceptable learning gains on these tests, but a normalized gain of ≥0.50 probably represents a substantial achievement.

In the course we taught, we devoted six 50-min lectures to teaching natural selection. This might seem excessive. Two lines of evidence suggest this time was well spent. First, after five lectures on natural selection, many of our students were unable to recognize that suicide was an unlikely adaptation. This implies their understanding of natural selection was still developing or they were unable to draw upon their knowledge when that knowledge was relevant. Second, the average test score on the short essay questions we used to assess understanding of natural selection did not appear to level off. Student answers appeared to be still improving at the end of our series of exercises (Figure 2). In particular, students' understanding of the genetic basis of evolution seemed to be improving (Figure 2).

We will conclude this paper with a few comments on a controversial aspect of moving from traditional lectures to active learning. Most of the learning exercises described here require a fair amount of time to conduct. This is especially true for the discussions of dog breeding, human evolution, and peacock trains. An instructor using these activities for the first time may find he or she has less time to cover other material. This could be viewed as a reason not to use active learning—especially if instructors feel pressure to cover as much material as possible. We have two responses. First, the goal of teaching is not to cover as much as possible, but to teach as much as possible, and a large body of evidence shows that students learn more in courses that make extensive use of active learning (e.g., Hake, 1998; Knight and Wood, 2005; Freeman *et al.*, 2007). Second, natural selection is one of the most important concepts in biology, but it is difficult for students to learn. All evidence suggests many students need specialized and time-consuming instruction to learn natural selection well enough to avoid falling back upon glaringly inaccurate misconceptions. Instructors might cover more if they used traditional lectures to cover natural selection, but it is hard to imagine what other topics students might learn that would make up for not understanding the principle cause of evolution.

ACKNOWLEDGMENTS

REFERENCES

Abrams E, Southerland S (2010). The hows and whys of biological change: how learners neglect physical mechanisms in their search for meaning. Int J Sci Educ *23*, 1271–1281.

Anderson DL, Fisher KM, Norman GJ (2002). Development and evaluation of the conceptual inventory of natural science. J Res Sci Teach *39*, 952–978.

Andrews TM, Kalinowski ST, Leonard MJ (2011a). "Are humans evolving?" A classroom discussion to change student misconceptions regarding natural selection. Evol Educ Outreach *4*, 456–466.

Andrews TM, Leonard MJ, Colgrove CA, Kalinowski ST (2011b). Active learning not associated with student learning in a random sample of college biology courses. CBE Life Sci Educ *10*, 394–405.

Baquero F, Blazquez J (1997). Evolution of antibiotic resistance. Trends Ecol Evol *12*, 482–487.

Barnett SM, Ceci SJ (2002). When and where do we apply what we learn?: A taxonomy for far transfer. Psychol Bull *128*, 612–637.

Bishop B, Anderson C (1990). Student conceptions of natural selection and its role in evolution. J Res Sci Teach *27*, 415–427.

Bransford JD, Brown AL, Cocking RR (eds.) (2000). How People Learn: Brain, Mind, Experience, and School, Washington, DC: National Academies Press.

Bransford JD, Schwartz D (1999). Rethinking transfer: a simple proposal with multiple implications. In: Review of Research in Education, vol. 24, ed. A Iran-Nejad and PD Pearson, Washington, DC: American Educational Research Association, 61–100.

Brumby MN (1984). Misconceptions about the concept of natural selection by medical biology students. Sci Educ *68*, 493–503.

Catrambone R, Holyoak KJ (1989). Overcoming contextual limitations on problem-solving transfer. J Exp Psychol *15*, 1147–1156.

Coyne J (2009). Why Evolution Is True, New York: Viking.

Darwin C (1859). On the Origin of Species, London: John Murray.

diSessa AA (2006). A history of conceptual change research: threads and fault lines. In: The Cambridge Handbook of the Learning Sciences, ed. RK Sawyer, New York: Cambridge University Press, 265–281.

Dobzhansky T (1973). Nothing in biology makes sense except in the light of evolution. Am Biol Teach 35, 125–12.

Duit R, Treagust DF (2003). Conceptual change: A powerful framework for improving science teaching and learning. Int J Sci Educ 25, 671–688.

Fisher K, Williams KS, Lineback JE, Anderson D (in prep.). Conceptual Inventory of Natural Selection—Abbreviated (CINS-abbr).

Ebert-May D, Derting TL, Hodder J, Momsen JL, Long TM, Jardeleza SE (2011). What we say is not what we do: effective evaluation of faculty professional development programs. Bioscience 61, 550–558.

Freeman S et al. (2007). Prescribed active learning increases performance in introductory biology. CBE Life Sci Educ 6, 132–139.

Gregory TR (2009). Understanding natural selection: essential concepts and common misconceptions. Evol Educ Outreach 2, 156–175.

Haak DC, HilleRisLambers J, Pitre E, Freeman S (2011). Increased structure and active learning reduce the achievement gap in introductory biology. Science 332, 1213–1216.

Hake RR (1998). Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. Am J Phys 66, 64–74.

Hammer D (1996). Misconceptions or P-Prims: how may alternative perspectives of cognition structure influence instructional perceptions and intentions? J Learn Sci 5, 97–127.

Hewson PW, Beeth ME, Thorley NR (1998). Teaching for conceptual change. In: International Handbook of Science Education, ed. BJ Fraser and KG Tobin, London: Kluwer Academic, 199–218.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. Science 313, 101–104.

Kalinowski ST, Leonard MJ, Andrews TM (2010). Nothing in evolution makes sense except in the light of DNA. CBE Life Sci Educ 9, 87–97.

Kaufman DW (1974). Adaptive coloration in Peromyscus polionotus: experimental selection by owls. J Mammol 55, 271–283.

Knight JK, Wood WB (2005). Teaching more by lecturing less. Cell Biol Educ 4, 298–310.

Lamarck J (1809). Philosophie Zoologique. Paris. Translated by H Elliot as Zoological Philosophy, London: Macmillan, 1914; reprinted by University of Chicago Press, 1984.

Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2006). SAS for Mixed Models, 2nd ed., Cary, NC: SAS Institute.

Lobato J (2008). Research methods for alternative approaches to transfer: implications for design experiments. In: Handbook of Design Research Methods in Education, ed. AE Kelly, RA Lesh, and JY Baek, New York: Routledge, 167–194.

Loyau A, Gomez D, Moureau B, Thery M, Hart NS, Saint Jalme M, Bennett ATD, Sorci G (2007). Iridescent structurally based coloration of eyespots correlates with mating success in the peacock. Behav Ecol 18, 1123–1131.

Marton F (2006). Sameness and difference in transfer. J Learn Sci 15, 499–535.

Mason L (ed.) (2007). Bridging the cognitive and sociocultural approaches in research on conceptual change. Educ Psychol. 42 (special issue).

Mestre J (2003). Transfer of Learning: Issues and Research Agenda: Report of a Workshop Held at the National Science Foundation. www.nsf.gov/pubs/2003/nsf03212/nsf03212.pdf (accessed 17 July 2013).

Miller JD, Scott EC, Okamoto S (2006). Public acceptance of evolution. Science 313, 765–766.

Murphy PK, Mason L (2006). Changing knowledge and beliefs. In: Handbook of Educational Psychology, 2nd ed., ed. PA Alexander and PH Winne, Mahwah, NJ: Lawrence Erlbaum, 305–324.

Nehm RH, Beggrow EP, Opfer JE, Ha M (2012). Reasoning about natural Selection: diagnosing contextual competency using the ACORNS instrument. Am Biol Teacher 74, 92–98.

Nehm RH, Ha M (2010). Item feature effects in evolution assessment. J Res Sci Teach 48, 237–256.

Nehm RH, Reilly L (2007). Biology majors' knowledge and misconceptions of natural selection. Bioscience 57, 263–272.

Nehm RH, Schonfeld I (2008). Measuring knowledge of natural selection: a comparison of the CINS, and open-response instrument, and oral interview. J Res Sci Teach 45, 1131–1160.

Petrie M (1994). Improved growth and survival of peacocks with more elaborate trains. Nature 3, 598–599.

Petrie M, Halliday T (1994). Experimental and natural changes in the peacock's (Pavo cristatus) train can affect mating success. Behav Ecol Sociobiol 35, 213–221.

Petrie M, Halliday T, Sanders C (1991). Peahens prefer peacocks with elaborate trains. Anim Behav 41, 323–331.

Scott PH, Asoko HM, Driver RH (1991). Teaching for conceptual change: a review of strategies. In: Research in Physics Learning: Theoretical Issues and Empirical Studies, ed. R Duit, F Goldberg, and H Niedderer, Kiel, Germany: IPN, 310–329.

Sinatra GM, Brem SK, Evans EM (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. Evol Educ Outreach 1, 189–195.

Sutter NB et al. (2007). A single IGF1 allele is a major determinant of small size in dogs. Science 316, 112–115.

Tanner K, Allen D (2005). Approaches to biology teaching and learning: understanding the wrong answers—teaching toward conceptual change. Cell Biol Educ 4, 112–117.

Turpen C, Finkelstein ND (2009). Not all interactive engagement is the same: variations in physics professors' implementation of peer instruction. Phys Rev Spec Top Phys Educ Res 5, 020101.

Vosniadou S (2008). International Handbook of Research on Conceptual Change, New York: Routledge.

Weinreich DM, Delaney NF, DePristo MA, Hartl DL (2006). Darwinian evolution can follow only very few mutation paths to fitter proteins. Science 312, 111–114.