Development and Evaluation of the Conceptual Inventory of Natural Selection

Dianne L. Anderson,[1] Kathleen M. Fisher,[1] Gregory J. Norman[2]

[1]*Center for Research in Math and Science Education, San Diego State University,
6475 Alvarado Road, Suite 206, San Diego, California 92120*

[2]*PACE Projects, San Diego State University, 6386 Alvarado Court, Suite 224,
San Diego, California 92120*

Abstract: Natural selection as a mechanism of evolution is a central concept in biology; yet, most nonbiology-majors do not thoroughly understand the theory even after instruction. Many alternative conceptions on this topic have been identified, indicating that the job of the instructor is a difficult one. This article presents a new diagnostic test to assess students' understanding of natural selection. The test items are based on actual scientific studies of natural selection, whereas previous tests have employed hypothetical situations that were often misleading or oversimplified. The Conceptual Inventory of Natural Selection (CINS) is a 20-item multiple choice test that employs common alternative conceptions as distractors. An original 12-item version of the test was field-tested with 170 nonmajors in 6 classes and 43 biology majors in 1 class at 3 community colleges. The test scores of one subset of nonmajors ($n = 7$) were compared with the students' performances in semistructured interviews. There was a positive correlation between the test scores and the interview scores. The current 20-item version of the CINS was field-tested with 206 students in a nonmajors' general biology course. The face validity, internal validity, reliability, and readability of the CINS are discussed. Results indicate that the CINS will be a valuable tool for instructors. © 2002 Wiley Periodicals, Inc. J Res Sci Teach 39: 952–978, 2002

Natural selection is the principal mechanism of evolution, and the theory of evolution is of great importance as a unifying theory in biology education according to the National Science Standards (National Research Council, 1996). Yet, natural selection is misunderstood by many students. The litany of alternative conceptions regarding natural selection and evolution is long (Mayr, 1982; Clough & Driver, 1986; Good, Trowbridge, Demastes, Wandersee, Hafner, & Cummins, 1992; Scharmann & Harris, 1992; Cummins, Demastes & Hafner, 1994). Some studies

focus specifically on subtopics of evolution such as natural selection (e.g., Brumby, 1979; Lawson & Thompson, 1988; Greene, 1990; Jiménez-Aleixandre, 1992, 1996; Grant, Owen, & Clarke, 1996; Holdredge, 1999), adaptation and reproduction (Lucas, 1971; Clough & Wood-Robinson, 1985; Renner, Brumby, & Shepherd, 1981), and speciation (Coyne, 1996). The Platonic attitudes that prevailed among the population 100 years ago (Mayr, 1982) persist in many quarters today (Almquist & Cronin, 1988; Greene, 1990; Jackson, Doster, Meadows, & Wood, 1995). Alternative conceptions about evolution exist even among the well-educated, including medical students (Brumby, 1984) and physics doctoral students (Chan, 1998). Other sources focus on how to teach evolution (e.g., Jensen & Finley, 1995; Tabak & Reiser, 1997). The National Academy of Sciences's book, *Teaching about Evolution and the Nature of Science* (1998), is especially helpful in this area.

Alternative conceptions are ideas that differ from the corresponding scientific explanations. They are usually held by a significant proportion of students and are highly resistant to instruction. At the same time, these alternative ideas can serve as anchoring conceptions (Clement, Brown, & Zietman, 1989; Otero, 2000) from which to move to a scientific conception when suitable instructional strategies are developed. An assessment tool that identifies alternative conceptions of students is desirable for teachers who are striving to promote constructivist learning in their classrooms (von Glasersfeld, 1989; Christianson & Fisher, 1999; Mintzes, Wandersee, & Novak, 2000). Whereas experienced teachers and professors are well aware of students' conceptual difficulties, novice teachers are not. In addition, many teachers recognize the need to assess their students' naive understandings but do not do so because they lack the appropriate tools (Morrison & Lederman, 2000). Thus, a reliable test with excellent content validity is needed to meet a variety of educational needs.

In this article, we describe a more realistic and comprehensive test for assessing conceptual understanding of natural selection. The item contexts are based on actual evolutionary events being studied by scientists, such as the Galapagos finches. We describe the development and evaluation of three test versions and include the current 20-item version of the Conceptual Inventory of Natural Selection (CINS) based on the best items to emerge from our research.

Semantic Pitfalls

Creating clear, unambiguous test items about natural selection is more challenging than writing test questions in many other areas of biology. The English language is laden with popular uses of words that seriously limit our ability to speak clearly and unambiguously in this field. For example, a question may refer to the owl. Is it referring to a pet owl, or to the various owl species that occupy the local farmlands, or to a local population of great horned owls, or to the entire great horned owl species? Cues in the context are essential for disambiguation because the word *owl* can be and often is used in each of these ways. Rarely do we take the time and make the effort to speak precisely, as in "the individual owl," "the local owls," "the local great horned owl population," and "the great horned owl species." We are prone to taking linguistic shortcuts.

In addition, there is a tendency to speak in a variety of ways that can cause (and often reflect) confusions. Examples include but are not limited to speaking anthropomorphically (bacteria have "shown considerable *ingenuity* in developing resistance to antibiotics"), teleologically ("cacti developed tough skin because it was *needed* to minimize water loss"), and as if evolution progresses steadily toward an ideal end point ("humans are *higher* on the evolutionary scale than chimps") (Jungwirth, 1975; Halldén, 1988; Pedersen & Halldén, 1994).

Jungwirth (1975) demonstrated that experts, as well as novices, frequently employ these various speech patterns. Presumably, the experts are speaking metaphorically rather than literally,

but they can confuse themselves on occasion and certainly can confuse their students. Ghiselin (1967) suggested that the habits and mannerisms of speech that characterize our language may often be derived from covert yet fundamental, metaphysical differences. Mayr (1982) reinforced this point of view with his explication of Platonic assumptions that have prevailed in western society for thousands of years and that stand in contradiction to neo-Darwinian thought.

Value of the CINS

With these semantic difficulties in mind, we examined existing tests for conceptual understanding of evolution (Bishop & Anderson, 1986, 1990; Settlage & Odom, 1995) and found that they suffered, like so many textbooks, from being overly simplistic and abstract. Some test questions even ask students to predict future evolutionary events when given information about a situation (Bishop & Anderson, 1990). Yet, predicting evolution is something Gould (1994) insisted even evolutionary biologists cannot do. In addition, existing tests directly seek to assess the student's understanding of the process of natural selection without probing the student's understanding of the underlying ecological and genetic principles that set the stage for natural selection. We believe that determining what students know about topics such as biotic potential and genetic variation within a population is essential.

Students who read Jonathan Weiner's Pulitzer prize–winning book, *The Beak of the Finch* (1995), develop a deeper and more comprehensive understanding of the processes of evolution and natural selection than any students we have ever seen who rely solely on lectures and introductory textbooks. We have found that even nonmajors, including prospective elementary school teachers, could develop a reasonable understanding of evolution with this entrancing story about 20 years of research on the evolution of finches in the Galapagos Islands, especially when guided with appropriate questions, prompts, discussions, and quizzes. This and other observations convinced us that the key to comprehension of this complex domain lies in the details, and that understanding the process of natural selection is just as important as understanding the outcome. We felt that this perspective could also be useful in the design of a new test.

The test items were chosen to address understanding of the five facts and three inferences described by Mayr (1982) that recreate the logic of the theory of natural selection. In addition, we have included questions to probe students' understanding of both the origin of variation and the origin of species. In all, a total of 10 concepts related to the theory of natural selection are represented on the CINS with 2 questions for each concept. The distractors in each item address common alternative conceptions about natural selection.

We are not aware of any existing test that is structured in this way, and believe that this type of test will be valuable in identifying gaps in instruction and student knowledge that contribute to misunderstandings about how populations change over time. We believe the test is best suited for assessing pre- and postinstruction knowledge of nonmajors and preinstruction knowledge of majors. We chose not to compare student performance on the CINS with performance on existing tests because the CINS is fundamentally different. Whereas the other tests assess the process of natural selection itself, the CINS also addresses the students' understanding of the underlying concepts of genetics and ecology that provide a foundation for using natural selection as an explanatory theory.

Perhaps the most effective way to identify misconceptions is to interview students. Obviously, this is logistically impossible in large classes. The goal of this research was to produce a test that would elicit information about student conceptions that paralleled information obtained in interviews, but could be used efficiently with large classes. When results on the CINS are compared with interview results, there is a positive correlation. This suggests that instructors

could use the easily administered CINS with large groups of students to obtain valid and reliable information on how well students understand the concepts. We hope that the CINS will be a useful tool in moving the research focus in evolution education from identification of alternative conceptions to developing explanations of how large groups of students progress in their understanding over time.

Natural Selection: Key Ideas Being Tested

The 20-item CINS presented in Appendix B assesses 10 main ideas. Italicized words in the following description of natural selection indicate all the ideas that appear in the test, and underlining indicates 10 main ideas.

Organisms produce more young than can be sustained by available resources (*biotic potential, carrying capacity*). All members of a species compete with one another for resources (*resources are limited*, *competition*), and some organisms do not survive (*limited survival*). Organisms within a species differ from one another in inherited traits (*genetic variation*). The variations arise through mutation and genetic recombination (*origin of variation*). Mutation and genetic recombination are *random* events that produce *beneficial*, *neutral*, *or harmful traits*. Much *variation is inherited* so the parents pass on their traits to their progeny. Among these offspring, those best suited to the environment tend to be most successful in producing young (*differential survival*, *fitness*, *reproductive success*). Offspring that are less well-suited to the environment are less likely to survive and less likely to produce offspring (*lowered fitness*, *lowered reproductive success*).

Through differential reproductive success, the frequency of different genetic types in the population can change with each succeeding generation (*descent with modification*, *evolution*, *change in gene pool over time*, *change in population*). Natural selection is *directed*, determined by the characteristics of the particular environment. Natural selection action on heritable traits is the primary mechanism of evolution. Action of natural selection on nonheritable traits has little long-term effect on the evolutionary process. Neither *disuse* of an organ nor *need* for a trait determine the genetic makeup of an organism. The *population* gradually becomes better suited to the environment through the propagation of more fit individuals (*adaptation*). Populations change through changes in the frequencies of genetic types in the population, not through change in individual organisms. This point is frequently missed by students.

When two populations of a single species are separated for an extended period of time by a physical, behavioral, temporal, or other barrier, the populations may diverge to the extent that they become separate species (*origin of species*). In general, members of different *species* are not able to mate and reproduce with one another, although this is not necessarily true of closely related species.

Test Design

In this section we describe the way that we developed the CINS. Our test items have authenticity by virtue of drawing on actual studies of evolution. This leads to a degree of precision and accuracy that is difficult to achieve in abstract and imaginary examples. The evolutionary setting is summarized in a heading, followed by related test items. Our examples are drawn from the abundance of microevolution studies cited in the literature [more than 500 in Endler (1986)]. We had hoped to include a plant example but did not locate one suitable for this context. The steps involved in test development are summarized briefly below. Our approach was similar but not identical to the method advocated by Treagust (1988).

1. Assess nonmajor student responses to open-ended test items about natural selection. A graduate student, C. Sandifer, and Fisher created an open-ended test about natural selection adapted from the Bishop and Anderson (1990) test and administered it to undergraduate non–biology majors. This provided a rich source of student beliefs about natural selection. Tamir (1971) first introduced this strategy as a means for identifying alternative conceptions and demonstrated the advantage of using alternative conceptions as distractors in multiple choice tests. Linke and Venz (1978), Halloun and Hestenes (1985), and Treagust (1988) used similar approaches.

2. Examine related literature. The authors of this article worked with an undergraduate (C. Goessling) and graduate students (L. Becvar and C. Noland) to find, read, and discuss relevant articles from the literature. The alternative conceptions described in the literature generally included and extended those displayed by the students in (Step 1) above.

3. Determine whether undergraduate biology majors achieve a successful understanding of natural selection. The difficulty of developing an accurate understanding of natural selection among non–biology majors is well documented. The researchers wanted to know the extent to which biology majors exhibited similar problems in understanding. A biology graduate student (A. Anderson) and Fisher collaboratively interviewed seven upper division ecology majors at a 4-year university. The interviews focused on natural selection, were semistructured, and were recorded on audiotape. A. Anderson transcribed the most interesting portions of the tapes (those in which students displayed difficulties). These seven interviewees demonstrated an excellent understanding of natural selection. We tried, largely unsuccessfully, to find weaknesses in their mental models. The students' weakest point was their understanding of the sources of genetic variation, but even there they knew the scientifically correct responses. They were simply less certain about their replies than in other areas. This suggests that evolution can be successfully taught to undergraduates when they have receptive worldviews and they receive significant exposure to the topic.

4. Define the content to be tested. The clear explications of natural selection by Mayr (1982), Endler (1986), and Jiménez-Aleixandre (1992, 1996) served as a basis for designing the test content. These were complemented by reading Malthus (1798/ 1971), Darwin (1859/1978), and Dobzhansky (1973). Mayr (1982) dissected the logic of the theory of natural selection into five facts from which three inferences are drawn.
Fact 1: All populations have the potential to grow at an exponential rate.
Fact 2: Most populations reach a certain size, then remain fairly stable over time.
Fact 3: Natural resources are limited.
       Inference 1: Not all offspring survive to reproductive age in part because of competition for natural resources.
Fact 4: Individuals in a population are not identical, but vary in many characteristics.
Fact 5: Many of the characteristics are inherited.
       Inference 2: Survival is not random. Those individuals with characteristics that provide them with some advantage over others in that particular environmental situation will survive to reproduce, whereas others will die.
       Inference 3: Populations change over time as the frequency of advantageous alleles increases. These could accumulate over time to result in speciation.

    In addition to these five facts and three inferences, we decided to include additional test items concerning the origin of variation and the origin of species. Accordingly, the final version of CINS includes 20 items, 2 for every concept.

5. Choose a style for the diagnostic test. Initial efforts focused on creating two-tiered test items similar to those employed by Treagust (1988), Odom and Barrow (1995), and

Settlage and Odom (1995). This test style is attractive because it separates factual knowledge (Tier 1 = facts) from reasons for choosing a particular fact (Tier 2 = mechanisms and beliefs).

Two-tiered questions work reasonably well for simple concepts such as the diffusion of atoms and molecules or the rules of Mendelian genetics, although even in such contexts they are not always as successful as one would hope (Griffard & Wandersee, 2001). However, with large, complex, multifaceted concepts such as evolution or natural selection, the two-tiered system breaks down. It is relatively easy to generate two or three answer choices to a content question in Tier 1, where one response is correct and the other one or two are plausible but incorrect. It then becomes extremely difficult to generate reasons for choosing one of the Tier 1 items that are plausible and also potentially applicable to more than one of those items. The Tier 2 responses we created generally (especially if correct) flagged one and only one Tier 1 response, and vice versa. We were unable to adapt this format to evolution because of the complexities involved. It would likely work well in a computer-administered test in which Tier 2 could have different choices for each different response in Tier 1. The simple multiple choice format of the CINS (as opposed to a computer format two-tier test) allows for easy administration to large groups of students as part of an in-class activity.

Instead, we followed Jungwirth's (1975) style, designing responses that aim to distinguish between different basic assumptions about the nature of the universe. This style involves use of relatively long text in both question header and item responses. The format does increase the authenticity and accuracy of the test, but at the same time we are concerned that it decreases readability. Thus, we include results of a readability test as part of Experiment 2.

6. Develop and validate a diagnostic test. The authors worked with Becvar, Noland, and Goessling to identify suitable evolutionary studies and develop related test items. The steps taken to develop and evaluate the items are described in the description of Experiments 1, 2, and 3 that follow.

7. Develop a specification grid. The specification grid (see Table 4 later) summarizes the scientific concepts on the 20-item test, the alternative conceptions used in distractors, and the item response choices that reflect each idea.

Experiment 1: Field Testing of CINS Version 1

Methods Used in Experiment 1

Approximately 100 students in four separate groups at large, ethnically diverse community colleges in Southern California participated in this study. All of the students were enrolled in nonmajors' general biology courses during the summer of 1999. Many students were preallied health majors. The students were given the tests as part of a regular, ungraded class activity that lasted approximately 15 minutes.

The version of the test used for this study included four situations: the Galapagos finches (Grant, Grant, & Petren, 2000, 2001; Petren, Grant, & Grant, 1999; Schluter, 2000; Weiner, 1995), the so-called blind cave salamanders, *Proteus anguinus*, in the Karst region of Europe (Culver, 1982), the peppered moth story *(Biston betularia)* of Great Britain's industrial revolution including aspects of both Kettlewell's (1955) original interpretation and the modern view (e.g., Grant et al., 1996), and the Canary Island lizards (Thorpe & Brown, 1989) that are believed to have migrated from Africa. Whereas the finch, lizard, and salamander examples illustrate natural selection leading to speciation, the moth example illustrates natural selection on a smaller scale.

The original CINS was composed of four sets of five questions in each section. Each group of students took two of the question sets (10 questions) as a pretest, and the other two sets (10 questions) as a posttest. Owing to the short period of time (approximately 10 days) between pre- and posttesting, using the same questions for both testing sessions could have been heavily biased as a result of priming effects. Because of this, each group of students was randomly split in half and given either finches/moths or lizards/salamanders as a pretest, and visa versa for the posttest.

As a pilot project, seven volunteers were recruited from one class to be interviewed about their understanding of natural selection both before and after instruction. Individual interviews lasted approximately 20 minutes and took place in a faculty office near the lecture room. The purpose of the interviews was to determine whether the students' scores on the CINS items were reflective of their understanding of the concepts as elicited during a one-on-one interview. The potential value of the test questions for classroom use depends on the ability of test results to predict the outcome of a one-on-one interview on the same topic. Audio recordings and extensive notes were made during each interview, and the pre- and postinterviews were transcribed.

Preinstruction interviews involved four tasks: (a) giving definitions of selected terms, (b) sorting cards into piles of related terms, (c) interviews about instances (Demastes, Good, & Peebles, 1996), and (d) diagram interpretation as described in Appendix A.

Students participated in 5 hours of lecture instruction and 3 hours of laboratory experience before postinterviews were conducted. Lecture instruction was traditional, with definitions and examples presented to students. One 3-hour lab session dealt with population ecology and included a simulated predator/prey interaction. Fewer interview tasks were used in the postinterviews than in the preinterviews, so that students could be asked about three of their answer choices on the CINS. Students were asked to read the question out loud, then go through each answer choice and explain why that was, or was not, a good choice.

To determine the validity of all the CINS items, 3 university and 2 community college biology professors were asked to choose the correct answers for each question. In addition to choosing the intended answer on each question, (thereby validating the tests), they also gave some helpful feedback that was used to improve the questions.

Results: Analysis of Items in Experiment 1

Significant problems were identified with the salamander questions during student interviews. The story is complex because the blind salamanders are actually born sighted and even appear to retain their sight mechanisms after a membrane forms over the eye during development, making them effectively blind except for responsiveness to flashes of light. Owing to problems in conveying these ideas clearly in the space available, the decision was made to drop those questions from the analysis.

The remaining 15 items were analyzed. In total, 105 students took the five moth questions, 117 students took the five lizard questions, and 101 students took the five finch questions. Initially, test results from the three groups of students were pooled into pretest and posttest for each section of the test. Results were scored on ParSCORE forms to obtain a detailed report. Because posttest scores showed little improvement over pretest scores, pre- and posttest results were pooled to increase the statistical power of the analysis. This lack of improvement is not surprising given the short instructional duration. With a total possible score of 5 points, the mean score for the five moth questions was 2.54 [standard deviation (SD) 1.36], for the five lizard questions was 2.36 (SD 1.46), and for the five finch questions, 2.12 (SD 1.15) so students averaged close to 50% for all three sets.

In a criterion-referenced test such as the CINS, the main objective is to assess mastery of subtopics. The goal is to create items that discriminate between students who exhibit mastery of a concept from those who do not. Discriminability of the items was determined by calculating point biserial values for each item; these ranged from 0.24 to 0.67 (Table 1). The point biserial method of assessing discriminability determines the correlation between performance on a particular item and overall performance on the test. Because our calculations are based on small 5-item sets, the performance on 1 item greatly influences the overall score. This reduces the usefulness of the point biserial value, yet these scores still provide some information about the discriminability of each item (Kaplan & Saccuzzo, 1997). The closer the point biserial value of an item is to 1.00, the greater is the discriminating power. All of the questions except for Finch Question 5 discriminate fairly well between low and high scoring students.

Question difficulty, which indicates the percentage of students choosing the correct answer, ranged from 13% to 77%. Optimum item difficulty is generally halfway between all students choosing the correct answer (100%) and the chance that a student would choose the right answer by guessing (25% because these items have four choices) (Kaplan & Saccuzzo, 1997). Given these guidelines, optimum difficulty is around 63%, and items with difficulty values of between 30% and 70% are best able to provide information about the differences between students (Kaplan & Saccuzzo, 1997). Two of the finch questions (1 and 5) appeared to be difficult for students. These items were carefully scrutinized and modified in the second version of the CINS.

Results: Test/Interview Comparisons in Experiment 1

The interviews suffered from the same semantic difficulties as were encountered in the attempts to prepare the two-tiered items described previously. These were exacerbated by the facts that many of the interview tasks were too general and were not closely aligned with the test items. However, the interviews were valuable in determining whether each student was capable of using the theory of natural selection to explain various situations. In addition, even though the

Table 1
Analysis of CINS data from Experiment 1 pooled across sites

Item	No.	Discriminability (Point Biserial Value)	Difficulty
Lizard 1	117	0.56	36%
Lizard 2	117	0.57	61%
Lizard 3	117	0.59	50%
Lizard 4	117	0.62	48%
Lizard 5	117	0.65	41%
Moth 1	105	0.63	51%
Moth 2	105	0.38	77%
Moth 3	105	0.59	33%
Moth 4	105	0.54	44%
Moth 5	105	0.67	49%
Finch 1	101	0.61	28%[b]
Finch 2	101	0.65	46%
Finch 3	101	0.52	75%
Finch 4	101	0.48	52%
Finch 5	101	0.24[a]	13%[b]

[a]Point biserial values are below the desirable minimum of 0.30.
[b]Items appear to be particularly difficult for students.

number of interview subjects was small, these results indicated that the test items had potential and that they warranted further research and development.

The think-aloud protocols were especially useful in providing specific information for comparing student knowledge in interviews with student knowledge displayed on the written test. Students who scored highly on the written test consistently performed well on the interview tasks. On the other hand, some students who appeared to have a relatively solid understanding of the concepts in interviews did not score highly on the test. This may be due in part to readability and also to the increased levels of discrimination required in the written items. In addition, it was obvious that some students had trouble with terminology. Words such as *characteristic* and *dominant* were unfamiliar to them in the context of heredity, and left these students confused.

The postinterview transcripts for each student were analyzed to determine the number of both scientifically correct utterances and utterances suggesting an alternative conception. The percentage of scientifically correct utterances out of the total was calculated. As shown in Table 2, there is a positive correlation between the two scores.

Two interview excerpts provide specific examples of the correlation between test scores and interview performance. Mark (a pseudonym) scored 9 of 10 on the posttest and provided many accurate explanations during the postinterview. In the first excerpt, Mark expresses an understanding that successful organisms pass their traits on to offspring while others do not.

> Interviewer: Define the term *adaptation* for me.
> Mark: Those are the traits that change over a period of time due to environment.
> Interviewer: So the environment causes the change?
> Mark: Yeah, the environment determines what is adaptable.
> Interviewer: So does [the environment] cause the changes that are necessary or does it act on things that are already there?
> Mark: I guess it acts on things already there. Because obviously, the animal that is more successful will pass on more of its genes, then over the course of time, those traits are going to be manifested in the offspring.

Later in the interview, Mark was shown a picture of a bird prying apart a pinecone and was asked to talk about how the picture illustrated the concept of natural selection.

> Interviewer: Have you ever tried to pull a pinecone apart?
> Mark: No, but I imagine it's probably a pretty hard thing to do. Because the dry ones are impossible, and that's a green one. So . . . "survival" and "adaptation," and I say "adaptation" because his beak and claws have probably adapted to doing this kind of thing, since it seems like he does it pretty easily.

Table 2

Relationship between participants' interview scores and CINS scores

Participant	Interview Score	CINS Score
1	44%	40%
2	46%	40%
3	53%	40%
4	71%	60%
5	61%	70%
6	65%	80%
7	85%	90%

Interviewer: Now when you say "adapted," do you think that individual bird changed in order to eat the pinecone?

Mark: Not this individual, I would say maybe his species probably evolved over time to be able to do this kind of thing based on availability of food, or whatever, maybe his species is the only one that can eat that particular kind of food because he is adapted to be able to do that. I guess also maybe his coloring is an adaptation. Maybe that is what helps them to recognize each other, make him more recognizable to females.

In contrast to Mark, Kim (a pseudonym) chose only 4 of 10 correct answers on the posttest. The following excerpts illustrate her basic misunderstanding of how natural selection influences populations. In the first excerpt, Kim was asked to think aloud as she read the second finch question of the CINS.

Kim: I think "A" would be better because it said "the need to be able to eat different foods," so if they were not able to eat those foods, they would die. So they would have to learn, I guess really quick, they had to change to be able to eat. And their beaks change.

Interviewer: So because they needed to, their beaks changed?

Kim: They *need* to eat, is like, if you don't, you die. *Want* is okay. I want to eat that, but if I don't, maybe I will get it next week or something. So it was a need, it was essential. I think that is the key word there– "need."

As Kim worked through another item (the third finch question on the CINS), she expresses some confused notions of how populations adapt.

Kim: "A" says that the finches were quite variable which evolution shows that everything usually comes from one thing. If evolution was changing their beaks, so how could they start at being quite variable? So it doesn't make sense. And usually [mumble] natural selection and stuff then you become suitable to get food, but if you were already like that, then there was no reason to change. It said "different lines of finches tried to develop different beaks," that's the same thing "evolution had to bring about the different lines" . . . so I didn't like that.

Mark is an older student who was taking general biology because of personal interest. His interviews demonstrated a working knowledge of many ideas related to natural selection, and his test score was high. Kim was a younger student with a minimal background in biology from high school. She was taking the course for general education units. Her interview demonstrated a great deal of confusion on the main ideas and her test score was poor. These interview results provided a glimpse into the potential usefulness of the test items, and we were encouraged to do more field testing as described in Experiment 2.

Experiment 2: Field Testing of CINS Version 2

Methods Used in Experiment 2

After the field testing and student interviews with Version 1 of CINS, some test items were revised based on student feedback from interviews and comments from biology professors. In addition, we replaced the salamander questions with a new set of questions based on natural selection experiments done with Venezualan guppies, *Poecilia reticulata* (Endler, 1980).

Segments of the inventory (i.e., all 5 of the finch questions and all 5 of the moth questions) were administered to students at a large urban community college in southern California as part of their regular class activities during the Fall 1999 semester. In some cases, student responses were collected after instruction on natural selection, and in some cases before instruction. Respondents included both nonmajors ($n = 53$) and majors ($n = 43$).

Results: Analysis of Items in Experiment 2

Discriminability (point biserial values) and difficulty were determined for all of the test items using a ParSCORE system, using data in groups of five items each (based on topic). These data are not reported here because of the questionable nature of using point biserial values based on a test composed of 5-item subsets. Even so, these data helped us to choose 12 items from the original pool of 20 items for inclusion in the next version of CINS. These 12 items target six of the main components of the theory of natural selection listed earlier (Mayr, 1982). These were chosen in part because we felt that the reliability of the CINS would be enhanced by having 2 items targeting each main idea. Point biserial values for the 12 items ranged from 0.14 to 0.62 and the difficulty on the items ranged from 18.9% to 84.1% correct.

Results: Readability of Question Stems in Experiment 2

After selection of 12 items for the inventory, the readability of the question stems was assessed using a rational deletion version of the cloze test (Taylor, 1953; Davies, 1995). In a cloze test, words are deleted from a passage, then a student is asked to complete the sentences. In this case, approximately every seventh word was deleted, but some selection was done to avoid deleting key, highly specific words that would be impossible for students to fill-in. A total of 15–23 community college students completed the cloze test for each question stem (Table 3). The percentage of exact matches was calculated for all question headers (lizard, finch, and moth) based on responses. If a student used a synonym for the intended term, it was counted as correct. A score of about 60% correct indicates a reading passage that is highly readable for the target group; a score of 40–60% is considered instructional level, indicating that the students could read the content with some support (P. Ross, personal communication, October 2000). The readability results for the question headers are shown in Table 3. They indicate that the item headers are at the appropriate reading level for the target audience.

Experiment 3: Field Testing of CINS Version 3

Methods Used in Experiment 3

After the field testing of CINS Version 2, new items were added to expand the CINS to include a total of 10 concepts related to natural selection. In addition, minor changes were made to

Table 3
Experiment 2 readability results for question headers

Question Header	Students (n)	Range of Scores	Mean Score
Moth	15	8–83%	53%
Lizard	17	15–85%	56%
Finch	23	17–72%	43%

improve the readability of some of the existing items. The items on CINS Version 3 address all 10 concepts twice as listed in Table 4. Items 1–10 include an item on each concept, as do Items 11–20 (although not in the same order). Table 4 also lists the alternative conceptions used in the distractors, and identifies the location of each distractor on the test. Finches are used as a context for Items 1–8, guppies are used in Items 9–13, and lizards are used in Items 14–20.

The entire 20-item CINS was administered to 206 students enrolled in two sections (A and B) of nonmajors' general biology course at a large urban community college in southern California during the Spring 2002 semester. The CINS was used as an in-class pretest before instruction on any topics related to natural selection. Students were given 30 minutes to complete the test and they received extra credit points based on the number of correct answers. The extra credit was offered to increase student motivation to answer the questions thoughtfully.

Results: Analysis of Items in Experiment 3

The 110 students in Section A of the course earned a mean score of 8.21 of a possible 20 items, with a range of 1–16 and an SD of 3.07. The 96 students in Section B earned a mean score of 10.42, with a range of 3–20 and an SD of 3.31. The demographics (ethnic diversity and male/female ratio) of the two sections were nearly identical, so the difference in mean scores may be due to the time constraints (another class waiting to enter the lecture hall) experienced by the students in Section A that did not affect the students in Section B. This explanation is supported by the fact that several students in Section A did not answer at least five of the questions, whereas only two students did not finish the CINS in Section B.

Discriminability (point biserial values) and difficulty values were determined for all of the test items using a ParSCORE system. Results for the two sections were averaged as shown in Table 5. The point biserial values indicate the ability of an individual item to discriminate between high and low performers on the entire test. The closer the point biserial value is to 1.00, the greater is the discriminating power. Good test items generally result in point biserial values of between 0.30 and 0.70 (Kaplan & Saccuzzo, 1997). Items 3, 4, 9–11, and 13 are below 0.30. Of these six, only Items 4 and 9 are significantly below the suggested minimum value. However, because the CINS is a criterion-referenced test designed to identify concepts that students do or do not understand, and not to discriminate among students, the point biserial values are of decreased usefulness (Gronlund, 1993). The difficulty of test items ranged from 14.5% to 80.6% of the students choosing the correct answer. The average difficulty was 46.4%, which is close to the 50% average difficulty suggested for a typical classroom test by Gronlund (1993).

The reliability of the test relates to the consistency of responses and the CINS must be shown to be reliable to be a valuable tool. As a measure of general internal consistency, we used the Kuder-Richardson 20. This method simultaneously considers all possible ways of splitting the test, so it improves on other methods of determining reliability in which the test is used only once (as opposed to test/retest methods). The KR_{20} for the test was 0.58 for Section A and 0.64 for Section B. A good classroom test should have a reliability coefficient of 0.60 or higher (Gronlund, 1993), so the CINS values are acceptable.

We examined the internal validity of the CINS with principal components analysis (PCA). This is a type of data reduction procedure that uses the covariance matrix of a set of p items to determine whether a smaller set of m components ($m < p$) can adequately explain the variation among the p items. Supporting evidence for the internal validity of a survey's underlying measurement structure is demonstrated if items measuring the same concept covary highly with each other and load on the same component, whereas items not measuring that same concept load on other components (DeVellis, 1991). A hypothesized underlying measurement structure of the

Table 4

Scientific concepts and alternative conceptions addressed in CINS Version 3

Topic	Scientific Concept	Alternative Conception
Biotic potential	All species have such great potential fertility that their population size would increase exponentially if all individuals that are born would again reproduce successfully (1C, 11B)	a) Not all organisms can achieve exponential population growth (11C) b) Organisms only replace themselves (1A, 11A) c) Populations level off (1B, 11D, 1D)
Population stability	Most populations are normally stable in size except for seasonal fluctuations (3B, 12A)	a) All populations grow in size over time (3A, 12B) b) Populations decrease (3D, 12C) c) Populations always fluctuate widely/randomly (3C, 12D)
Natural resources	Natural resources are limited; nutrients, water, oxygen, etc. necessary for living organisms are limited in supply at any given time (2A, 14D)	Organisms can always obtain what they need to survive (2B, 2C, 2D, 14A, 14B, 14C)
Limited survival	Production of more individuals than the environment can support leads to a struggle for existence among individuals of a population, with only a fraction surviving each generation (5D, 15D)	a) There is often physical fighting among one species (or among different species) and the strongest ones win (5B, 15B) b) Organisms work together (cooperate) and don't compete (5A, 5C, 15A)
Variation within a population	Individuals of a population vary extensively in their characteristics (9D, 16C)	a) All members of a population are nearly identical (9A, 16A) b) Variations only affect outward appearance, don't influence survival (9B, 9C, 16B) c) Organisms in a population share no characteristics with others (16D)
Variation inheritable	Much variation is heritable (7C, 17D)	a) When a trait (organ) is no longer beneficial for survival, the offspring will not inherit the trait (7B, 17B) b) Traits acquired during an organism's lifetime will be inherited by offspring (7A, 17A) c) Traits that are positively influenced by the environment will be inherited by offspring (7D)
Differential survival	Survival in the struggle for existence is not random, but depends in part on the hereditary constitution of the surviving individuals. Those individuals whose surviving characteristics fit them best to their environment are likely to leave more offspring than less fit individuals (10C, 18B)	a) Fitness is equated with strength, speed, intelligence or longevity (10A, 10B, 18A, 18C, 18D) b) Organisms with many mates are biologically fit (10D)

(*Continued*)

Table 4
(Continued)

Topic	Scientific Concept	Alternative Conception
Change in a population	The unequal ability of individuals to survive and reproduce will lead to gradual change in a population, with the proportion of individuals with favorable characteristics accumulating over the generations (4B, 13B)	a) Changes in a population occur through a gradual change in all members of a population (4A, 13A, 17C) b) Learned behaviors are inherited (4C, 13C) c) Mutations occur to meet the needs of the population (4D, 13D)
Origin of species	An isolated population may change so much over time that it becomes a new species (8A, 20B)	a) Organisms can intentionally become new species over time (an organism tries, wants, or needs to become a new species) (8C, 8D, 20A, 20D) b) Speciation is a hypothetical idea (8B, 20C)
Origin of variation	Random mutations and sexual reproduction produce variations; while many are harmful or of no consequence, a few are beneficial in some environments (6B, 19C)	a) Mutations are adaptive responses to specific environmental agents (6C, 15C, 19D) b) Mutations are intentional: an organism tries, needs, or wants to change genetically (6A, 6D, 19A, 19B)

Table 5
Analysis of CINS data from Experiment 3 (average for two sections of students taking General Biology, n = 206)

Item	Concept	Discriminability (Point Biserial Value)	Difficulty (% Correct Responses)
1	Biotic potential	0.36	69.4%
11	Biotic potential	0.26[a]	63.1%
2	Pop. are stable	0.49	61.2%
14	Pop. are stable	0.43	51.5%
3	Resources limited	0.29[a]	48.7%
12	Resources limited	0.33	48.7%
5	Limited survival	0.47	67.2%
15	Limited survival	0.52	42.3%
9	Variation	0.20[a]	52.0%
16	Variation	0.34	80.6%
7	Variation inherited	0.40	55.0%
17	Variation inherited	0.34	38.8%
10	Differential survival	0.27[a]	55.5%
18	Differential survival	0.41	39.1%
4	Change in population	0.16[a]	18.2%[b]
13	Change in population	0.29[a]	28.3%
6	Origin of variation	0.39	14.5%[b]
19	Origin of variation	0.35	33.7%
8	Origin of species	0.32	41.4%
20	Origin of species	0.31	22.3%[b]

[a]Point biserial values are below the desirable minimum of 0.30.
[b]Items appear to be particularly difficult for students.

CINS would be to have 10 components explaining much of the variation among the 20 items. Each of the 10 components would represent a distinct evolutionary concept with the two items designed to measure a particular concept both loading on the appropriate component.

The PCA was conducted on the 20×20 matrix of item phi correlation coefficients. Two methods estimated the number of components to retain (Horn, 1965; Lautenschlager, 1989). Parallel analysis, which determines the number of components by comparing the eigen values of the observed correlation matrix with those from a matrix of the same variable and sample size based on random variables, indicated 2 components should be extracted. The number of eigenvalues >1 rule indicated 8 components should be extracted. The varimax rotated component patterns of Solutions 2–8 were examined. Criteria for determining the final PCA solution included: (a) having a large proportion of the total matrix variation explained, (b) having a high number of items with a strong ($>.40$) loading on at least one component, (c) having a minimum number of complex items (items with strong loadings on more than one component), and (d) having a component pattern that was theoretically interpretable.

The 7-component extraction was found to be optimal. Seven components accounted for 53% of the total variance. All 20 items had a loading $>.40$ on at least one component. Only one item (12) had a loading $>.40$ on multiple components (Components 3 and 5). Nine of the 10 pairs of items representing the 10 evolution concepts emerged together on the same component (Table 6). Only the paired items measuring Variation Inherited (Items 7 and 17) loaded on two different components. Two components contained multiple pairs of items. Component 1 included Natural Resources (Items 2 and 14) and Limited Survival (Items 5 and 15) and Component 4 contained Change in Population (Items 4 and 13) and Origin of Species (Items 8 and 20). Overall, the PCA results indicated support for the internal validity of the CINS Version 3 instrument.

Conclusions

The Bishop and Anderson (1990) test of understanding of natural selection has been extremely popular and has probably been used by many hundreds of teachers and researchers. We have made good use of it ourselves. So why develop another test? As noted in the introduction, we observe that when nonmajors read and discuss Jonathan Weiner's *Beak of the Finch*, they do acquire a much deeper understanding of evolution than they ever seem to obtain from abstract descriptions and/or short simulations. This led us to think that a test also would also be less ambiguous and more accurate if it focused on actual scientific data.

We were encouraged by our finding that the seven undergraduate biology majors whom we interviewed all demonstrated an accurate understanding of natural selection. This suggests to us that it is possible for students to learn about natural selection and that we should be much more successful than we currently are with nonmajors. We believe we have discovered the means for achieving this goal.

The *Beak of the Finch* is a Pulitzer prize–winning book that describes research into the microevolution of the Galapagos finches over a 20-year period. The book describes how the data are collected by scientists who camp out on this forlorn island for months at a time, year after year. Students can see the changes that occur in the overall group of finches as the island goes through first a multiyear drought and then a tumultuous rainy period. Large changes occur in the size of the finch population. Pronounced changes occur in the proportions of different finch species that make up the groups. Clearly measurable changes occur in the average length of the beaks in the populations of each finch species. Students also learn the ways in which the data are carefully organized into large computers back at Princeton and analyzed in many different ways. Going

Table 6
Principle components analysis with seven-component solution

Item	Component						
	1	2	3	4	5	6	7
Biotic potential							
1			.624				
11			.714				
Stable population							
3					.845		
12			.455		.596		
Natural resources							
2	.706						
14	.502						
Limited survival							
5	.569						
15	.589						
Variation							
9							.737
16							.547
Variation inherited							
7	.502						
17		.687					
Differential survival							
10						.769	
18						.472	
Change in population							
4				.406			
13				.671			
Origin of variation							
6		.501					
19		.659					
Origin of species							
8				.418			
20				.593			

through the actual research process and seeing the actual research results provides a tangible model for understanding the basic concepts.

If a book can provide this sort of detailed learning model, we predicted that a diagnostic test could be prepared that would present questions within realistic contexts. These contexts would help students to make sense of the information and to express their understanding (or misunderstanding) of the theory of natural selection as it applies to various examples by focusing on details of the changes.

The CINS (Appendix B) was designed to be used as a tool by teachers and professors interested in instructional methods that support constructivist and socioconstructivist learning. Our results support the prediction that a multiple choice test could be prepared using realistic topics as well as using common alternative conceptions as distractors. These results also indicate that the scores on the test correlate positively with scores on one-on-one interviews. Therefore, the CINS provides a simple yet effective means of identifying the frequency of some common misconceptions among large numbers of students. Because the inventory targets 10 of the main ideas of the

theory of natural selection, instructors can be assured that a student who chooses all of the correct answers has a fairly comprehensive understanding of how natural selection influences populations.

The field testing reported in this article indicates that the CINS's face validity has been verified by independent content experts. In addition, the readability of the question stems has been evaluated and is at a reasonable level for first-year college students. The reliability of the CINS as determined by the KR_{20} is acceptable. The discriminability and difficulty scores of the question items indicate that the questions are capable of distinguishing between students who understand the concepts and those who do not.

The results of the principle components analysis demonstrate strong support for the internal validity of the CINS's underlying measurement structure. The seven components represented distinct evolutionary concepts. Five of the components contained a single set of item pairs representing one concept. Components 1 and 4 revealed where items intended to measure different concepts were similar to each other. First, the items related to natural resources being limited and the items related to the idea that not all individuals survive were both contained in Component 1. This is not surprising because when students understand that there is competition for resources, they acknowledge that some individuals die. Second, the items related to a change in a population over time and the origin of species were both contained in Component 4. This pairing also is reasonable because a change in allele frequency of a population (Items 4 and 13) will eventually set the stage for the origin of new species (Items 8 and 20) as two populations become reproductively isolated in some way. The analysis also showed that the items designed to measure Variation Inherited related more to other items than they did to each other. We are not able to explain why Item 7 on how variation is inherited clustered with items on natural resources and limited survival as these ideas are distinct. However, Item 17 is related to the items on the origin of variation (Items 6 and 19). This points to where the measure can be further improved so that items related to the inheritance of variation (Items 7 and 17) will match more closely. One slight change has been made to the wording of a distractor on Item 17 to make it more clear. Further field testing will reveal whether this change causes Items 7 and 17 to pair as they should.

Although the interview data reported here are limited, the results indicate that a high score on the easily administered CINS correlates with a high degree of understanding of natural selection during an interview. For this reason, the CINS should be a useful instrument for investigating student conceptions with hundreds of students. This provides researchers with a means of going beyond studies with a small number of interviews that are difficult to generalize. The first author is currently conducting additional interviews with general biology students and comparing these interview results with performance on the CINS.

The CINS is similar in structure to the Project Star Astronomy Concept Inventory that assesses understanding of concepts with distractor-driven multiple choice items (Sadler, 1998). Sadler suggested that such test items have a very different item profile for students of varying abilities than items without such attractive distractors. In fact, Sadler reported that students may have certain inaccurate ideas strengthened before coming to a point of understanding a particular concept in a scientific way. For this reason, further research with the CINS may involve using Item Response Theory and item option characteristic curves (Sadler, 1998), rather than Classical Test Theory based on discriminability and difficulty, to evaluate the performance of large groups of students at varying levels of ability (high school, college nonmajors, and college biology majors).

In addition to being a useful student assessment, the CINS items are potentially valuable for in-class discussion. Occasional use of selected items in this way always generates lively and highly productive discussion with nonmajors. The questions serve as an easy to use format for

getting students to think about what they know, so that new information can be compared to prior knowledge, and the stage is set for conceptual change.

The theory of natural selection provides a unifying and explanatory framework for biology. As biology educators, we owe our nonmajors an opportunity to develop a working knowledge of the theory as a mechanism of evolution, because they will likely not have another chance to learn this fundamental idea. If this diagnostic test can be used to assess instructional methods or to stimulate conceptual change in these students, we will have met our goal in producing it.

Appendix A: Experiment 1 Interview Tasks

Task 1—definition of terms

Instructions:
''Please tell me your understanding of these terms.''

natural selection
population
fitness
mutation
species
competition

Task 2—card sort

Instructions:
''Please arrange these cards on the table so that the words that are closely related to each other are close together, and those that are unrelated to each other are far apart. If there are any terms that are unfamiliar to you or have nothing to do with natural selection, put them aside.''

adaptation	need
competition	offspring
fitness	population of rabbits
gene	random
individual rabbit	survival
mutation	variation
want	

Task 3—interview about instances

Instructions:
''I am going to show you several pictures now. Please tell me whether each one is an example of one or more of the terms on this card, then explain your answer.''

Terms on card:
competition
variation within a species
variation between species
survival

Photos shown during preinstruction interviews:

1. Lizard camouflaged against lichen-covered rock
2. Four distinctly different types of antelope
3. A hedgehog
4. A meerkat colony
5. Raccoons feeding on garbage from overflowing garbage cans
6. Dozens of flamingos standing on or near nests, some chicks visible
7. Tropical island foliage
8. Giraffe pulling leaves off a thorny acacia branch

Photos shown during postinstruction interviews:

1. Cheetah camouflaged in grasses
2. Four distinctly different monkey species
3. Deer pulling the last leaves off branches during a snowstorm
4. Hundreds of birds nesting on a rocky beach
5. Parrot prying open a pine cone with its beak

Interview task 4—butterfly population (used only in preinstruction interviews)

Instructions:

''Assume that these circles represent a population of butterflies in a population. Notice that one butterfly is dark, while all of the others are light. What could have made this butterfly, or its ancestors, dark?'' (The student answers the question.)
''If being dark somehow increased the chances of survival, what might happen to the population over a long period of time? Color in the circles to show your answer.''

Students were shown a diagram with three rows of circles representing 3 generations of a population. In the first generation, one circle was solid black; the others were open. All circles in the other two rows (generations) were open.

Appendix B: Conceptual Inventory of Natural Selection

Answer key: 1-C, 2-A, 3-B, 4-B, 5-D, 6-B, 7-C, 8-A, 9-D, 10-C, 11-B, 12-A, 13-B, 14-D, 15-D, 16-C, 17-D, 18-B, 19-C, 20-B.

Conceptual Inventory of Natural Selection

D.L. Anderson and K.M. Fisher

Your answers to these questions will assess your understanding of the Theory of Natural Selection. Please choose the answer that best reflects how a biologist would think about each question.

Galapagos finches
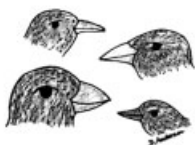
Scientists have long believed that the 14 species of finches on the Galapagos Islands evolved from a single species of finch that migrated to the islands one to five million years ago (Lack, 1940). Recent DNA analyses support the conclusion that all of the Galapagos finches evolved from the warbler finch (Grant, Grant & Petren, 2001; Petren, Grant & Grant,, 1999). Different species live on different islands. For example, the medium ground finch and the cactus finch live on one island. The large cactus finch occupies another island. One of the major changes in the finches is in their beak sizes and shapes, as shown in this figure.

Choose the one answer that best reflects how an evolutionary biologist would answer.

1. What would happen if a breeding pair of finches was placed on an island under ideal conditions with no predators and unlimited food so that all individuals survived? Given enough time
    a. the finch population would stay small because birds only have enough babies to replace themselves.
    b. the finch population would double and then stay relatively stable.
    c. the finch population would increase dramatically.
    d. the finch population would grow slowly and then level off.

2. Finches on the Galapagos Islands require food to eat and water to drink.
    a. When food and water are scarce, some birds may be unable to obtain what they need to survive.
    b. When food and water are limited, the finches will find other food sources, so there is always enough.
    c. When food and water are scarce, the finches all eat and drink less so that all birds survive.
    d. There is always plenty of food and water on the Galapagos Islands to meet the finches' needs.

3. Once a population of finches has lived on a particular island for many years,
    a. the population continues to grow rapidly.
    b. the population remains relatively stable, with some fluctuations.
    c. the population dramatically increases and decreases each year.
    d. the population will decrease steadily.

4. In the finch population, what are the primary changes that occur gradually over time?
    a. The traits of each finch within a population gradually change.
    b. The proportions of finches having different traits within a population change.
    c. Successful behaviors learned by finches are passed on to offspring.
    d. Mutations occur to meet the needs of the finches as the environment changes.

5. Depending on their beak size and shape, some finches get nectar from flowers, some eat grubs from bark, some eat small seeds, and some eat large nuts. Which statement best describes the interactions among the finches and the food supply?
    a. Most of the finches on an island cooperate to find food and share what they find.
    b. Many of the finches on an island fight with one another and the physically strongest ones win.
    c. There is more than enough food to meet all the finches' needs so they don't need to compete for food.
    d. Finches compete primarily with closely related finches that eat the same kinds of food, and some may die from lack of food.

6. How did the different beak types <u>first</u> arise in the Galapagos finches?
    a. The changes in the finches' beak size and shape occurred because of their need to be able to eat different kinds of food to survive.
    b. Changes in the finches' beaks occurred by chance, and when there was a good match between beak structure and available food, those birds had more offspring.
    c. The changes in the finches' beaks occurred because the environment induced the desired genetic changes.
    d. The finches' beaks changed a little bit in size and shape with each successive generation, some getting larger and some getting smaller.

7. What type of variation in finches is passed to the offspring?
    a. Any behaviors that were learned during a finch's lifetime
    b. Only characteristics that were beneficial during a finch's lifetime
    c. All characteristics that are genetically determined
    d. Any characteristics that were positively influenced by the environment during a finch's lifetime

8. What caused populations of birds having different beak shapes and sizes to become distinct species distributed on the various islands?
    a. The finches were quite variable, and those whose features were best suited to the available food supply on each island reproduced most successfully.
    b. All finches are essentially alike and there are <u>not</u> really fourteen different species.
    c. Different foods are available on different islands and for that reason, individual finches on each island gradually developed the beaks they needed.
    d. Different lines of finches developed different beak types because they needed them in order to obtain the available food.

Venezuelan Guppies

Guppies are small fish found in streams in Venezuela. Male guppies are brightly colored, with black, red, blue and iridescent (reflective) spots. Males cannot be too brightly colored or they will be seen and consumed by predators, but if they are too plain, females will choose other males. Natural selection and sexual selection push in opposite directions. When a guppy population lives in a stream in the absence of predators, the proportion of males that are bright and flashy increases in the population. If a few aggressive predators are added to the same stream, the proportion of bright-colored males decreases within about five months (3-4 generations). The effects of predators on guppy coloration have been studied in artificial ponds with mild, aggressive, and no predators, and by similar manipulations of natural stream environments (Endler, 1980).

Choose the one answer that best reflects how an evolutionary biologist would answer.

9. A typical natural population of guppies consists of hundreds of guppies. Which statement best describes the guppies of a single species in an isolated population?
    a. The guppies share all of the same characteristics and are identical to each other.
    b. The guppies share all of the essential characteristics of the species; the minor variations they display don't affect survival.
    c. The guppies are all identical on the inside, but have many differences in appearance.
    d. The guppies share many essential characteristics, but also vary in many features.

10. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Which feature would a biologist consider to be most important in determining which guppies were the "most fit"?
    a. large body size and ability to swim quickly away from predators
    b. excellent ability to compete for food
    c. high number of offspring that survived to reproductive age
    d. high number of matings with many different females.

11. Assuming ideal conditions with abundant food and space and no predators, what would happen if a pair of guppies were placed in a large pond?
    a. The guppy population would grow slowly, as guppies would have only the number of babies that are needed to replenish the population.
    b. The guppy population would grow slowly at first, then would grow rapidly, and thousands of guppies would fill the pond.
    c. The guppy population would never become very large, because only organisms such as insects and bacteria reproduce in that manner.
    d. The guppy population would continue to grow slowly over time.

12. Once a population of guppies has been established for a number of years in a real (not ideal) pond with other organisms including predators, what will likely happen to the population?
    a. The guppy population will stay about the same size.
    b. The guppy population will continue to rapidly grow in size.
    c. The guppy population will gradually decrease until no more guppies are left.
    d. It is impossible to tell because populations do not follow patterns.

13. In guppy populations, what are the primary changes that occur gradually over time?
    a. The traits of each individual guppy within a population gradually change.
    b. The proportions of guppies having different traits within a population change.
    c. Successful behaviors learned by certain guppies are passed on to offspring.
    d. Mutations occur to meet the needs of the guppies as the environment changes.

Canary Island Lizards
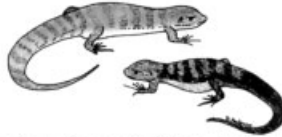
The Canary Islands are seven islands just west of the African continent. The islands gradually became colonized with life: plants, lizards, birds, etc. Three different species of lizards found on the islands are similar to one species found on the African continent (Thorpe & Brown, 1989). Because of this, scientists assume that the lizards traveled from Africa to the Canary Islands by floating on tree trunks washed out to sea.

Choose the one answer that best reflects how an evolutionary biologist would answer.

14. Lizards eat a variety of insects and plants. Which statement describes the availability of food for lizards on the Canary Islands?
    a. Finding food is not a problem since food is always in abundant supply.
    b. Since lizards can eat a variety of foods, there is likely to be enough food for all of the lizards at all times.
    c. Lizards can get by on very little food, so the food supply does not matter.
    d. It is likely that sometimes there is enough food, but at other times there is not enough food for all of the lizards.

15. What do you think happens among the lizards of a certain species when the food supply is limited?
    a. The lizards cooperate to find food and share what they find.
    b. The lizards fight for the available food and the strongest lizards kill the weaker ones.
    c. Genetic changes that would allow lizards to eat new food sources are likely to be induced.
    d. The lizards least successful in the competition for food are likely to die of starvation and malnutrition.

16. Populations of lizards are made up of hundreds of individual lizards. Which statement describes how similar they are likely to be to each other?
    a. All lizards in the population are likely to be nearly identical.
    b. All lizards in the population are identical to each other on the outside, but there are differences in their internal organs such as how they digest food.
    c. All lizards in the populations share many similarities, but there are differences in features like body size and claw length.
    d. All lizards in the population are completely unique and share no features with other lizards.

17. Which statement could describe how traits in lizards pass from one generation of lizards to the next generation?
    a. Lizards that learn to catch a particular type of insect will pass the new ability to offspring.
    b. Lizards that are able to hear, but have no survival advantage because of hearing, will eventually stop passing on the "hearing" trait.
    c. Lizards with stronger claws that allow for catching certain insects have offspring whose claws gradually get even stronger during their lifetime.
    d. Lizards with a particular coloration and pattern are likely to pass the same trait on to offspring.

18. Fitness is a term often used by biologists to explain the evolutionary success of certain organisms. Below are descriptions of four fictional female lizards. Which lizard might a biologist consider to be the "most fit"?

	Lizard A	Lizard B	Lizard C	Lizard D
Body length	20 cm	12 cm	10 cm	15 cm
Offspring surviving to adulthood	19	28	22	26
Age at death	4 years	5 years	4 years	6 years
Comments	Lizard A is very healthy, strong, and clever	Lizard B has mated with many lizards	Lizard C is dark colored and very quick	Lizard D has the largest territory of all the lizards

    a. Lizard A
    b. Lizard B
    c. Lizard C
    d. Lizard D

19. According to the theory of natural selection, where did the variations in body size in the three species of lizards most likely come from?
    a. The lizards needed to change in order to survive, so beneficial new traits developed.
    b. The lizards wanted to become different in size, so beneficial new traits gradually appeared in the population.
    c. Random genetic changes and sexual recombination both created new variations.
    d. The island environment caused genetic changes in the lizards.

20. What could cause one species to change into three species over time?
    a. Groups of lizards encountered different island environments so the lizards needed to become new species with different traits in order to survive.
    b. Groups of lizards must have been geographically isolated from other groups and random genetic changes must have accumulated in these lizard populations over time.
    c. There may be minor variations, but all lizards are essentially alike and all are members of a single species.
    d. In order to survive, different groups of lizards needed to adapt to the different islands, and so all organisms in each group gradually evolved to become a new lizard species.

References

Almquist, A.J. & Cronin, J.E. (1988). Fact, fancy and myth on human evolution. Current Anthropology, 29, 520–522.

Bishop, B.A. & Anderson, C.W. (1986). Evolution by teaching natural selection: A teaching module. Occasional paper 91. Institute for Research on Teaching, Michigan State University, East Lansing, MI.

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

Brumby, M.N. (1979). Students' perceptions and learning styles associated with the concept of evolution by natural selection. Unpublished doctoral dissertation, University of Surrey, United Kingdom.

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

Chan, K.S. (1998, April). A case study of a physicist's conceptions about the theory of evolution. Paper presented at the annual meeting of the National Association of Research and Science Teaching, San Diego, CA.

Christianson, R.G. & Fisher, K.M. (1999). Comparison of student learning about diffusion and osmosis in constructivist and traditional classrooms. International Journal of Science Education, 21, 687–698.

Clement, J., Brown, D.E., & Zietman, A. (1989). Not all preconceptions are misconceptions: Finding ''anchoring conceptions'' for grounding instruction on students' intuitions. International Journal of Science Education, 11, 554–565.

Clough, E.E. & Driver, R. (1986). A study of the consistency in the use of students' conceptual frameworks across different task contexts. Science Education, 70, 473–496.

Clough, E.E. & Wood-Robinson, C. (1985). How secondary students interpret instances of biological adaptation. Journal of Biological Education, 19, 304–310.

Coyne, J. (1996). Speciation in action. Science, 272, 700–701.

Culver, D.C. (1982). Cave life: Evolution and ecology. Cambridge, MA: Harvard University Press.

Cummins, C.L., Demastes, S.S., & Hafner, M.S. (1994). Evolution: Biology education's under-researched unifying theme. Journal of Research in Science Teaching, 31, 445–448.

Davies, F. (1995). Introducing reading. London: Penguin English.

Darwin, C. (1859/1978). The origin of species. New York: Penguin Books. (Original work published 1859).

Demastes, S.S., Good, R.G., & Peebles, P. (1996) Patterns of conceptual change in evolution. Journal of Research in Science Teaching, 33, 407–431.

DeVellis, R.F. (1991). Scale development. Newbury Park, CA: Sage.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. American Biology Teacher, 35, 125–129.

Endler, J.A. (1980). Natural selection on color patterns in *Poecilia reticulata*. Evolution, 34, 76–91.

Endler, J.A. (1986). Natural selection in the wild. Princeton: Princeton University Press.

Ghiselin, M.T. (1967). On semantic pitfalls of biological adaptation. Philosophy of Science, 34, 147–153.

Good, R.G., Trowbridge, J.E., Demastes, S.S., Wandersee, J.H., Hafner, M.S., & Cummins, C.L. (1992, December). Toward a research base for evolution education: Report of a national conference. EDRS Conference Proceedings, ED 361 183, SE 053 585, Evolution Education Research Conference, Baton Rouge, LA.

Gould, S.J. (1994). The evolution of life on earth. Scientific American, October, 85–91.

Grant, B.S., Owen, D.F., & Clarke, C.A. (1996). Parallel rise and fall of melanic peppered moths in America and Britain. Journal of Heredity, 87, 351–357.

Grant, P.R., Grant, B.R., & Petren, K. (2000). The allopatric phase of speciation: The sharp-beaked ground finch *(Geospiza difficilis)* on the Galapagos islands. Biological Journal of the Linnean Society, 69, 287–317.

Grant, P.R., Grant, B.R., & Petren, K. (2001). A population founded by a single pair of individuals: Establishment, expansion, and evolution. Genetica, 112–113, 359–382.

Greene, E.D., Jr. (1990). The logic of university students' understanding of natural selection. Journal of Research in Science Teaching, 27, 875–885.

Griffard, P.B. & Wandersee, J.H. (2001, March). A qualitative look at a quantitative approach to alternative conceptions research: The two-tier instrument. Presented at the annual meeting of the National Association for Research in Science Teaching, St. Louis, MO.

Gronlund, N.E. (1993) How to make achievement tests and assessments (5th ed) Boston: Allyn & Bacon.

Halldén, O. (1988). The evolution of the species: Pupil perspectives and school perspectives. International Journal of Science Education, 10, 541–552.

Halloun, I.A. & Hestenes, D. (1985). The initial knowledge of college physics students. American Journal of Physics, 53, 1043–1055.

Holdredge, C. (1999). The case of the peppered moth illusion. Whole Earth, 66–69.

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 179–185.

Jackson, D.F., Doster, E.C., Meadows, L., & Wood, T. (1995). Hearts and minds in the science classroom: The education of a confirmed evolutionist. Journal of Research in Science Teaching, 32, 585–611.

Jensen, M.S. & Finley, F.N. (1995). Teaching evolution using historical arguments in a conceptual change strategy. Science Education, 79, 147–166.

Jiménez-Aleixandre, M.P. (1992). Thinking about theories or thinking with theories: A classroom study with natural selection. International Journal of Science Education, 14, 51–61.

Jiménez-Aleixandre, M.P. (1996). Darwinian and Lamarckian models used by students and their representations. In Fisher, K.M. & Kibby, M. (Eds.), Knowledge acquisition, organization and use in biology (pp 65–77). New York: Springer Verlag.

Jungwirth, E. (1975). Preconceived adaptation and inverted evolution. Australian Science Teachers Journal, 21, 95–100.

Kaplan, R.M. & Saccuzzo, D.P. (1997). Psychological testing: Principles, applications, and issues (4th ed.) Pacific Grove, CA: Brooks/Cole.

Kettlewell, H.B.D. (1955). Selection experiments on industrial melanism in the Lepidoptera. Heredity, 9, 323–342.

Lack, D. (1940). Evolution of the Galapagos finches. Nature, 146, 324–327.

Lautenschlager, G.J. (1989). A comparison of alternatives to conducting Monte Carlo analyses for determining parallel analysis criteria. Multivariate Behavioral Research, 24, 365–395.

Lawson, A.E. & Thompson, L.D. (1988). Formal reasoning ability and misconceptions concerning genetics and natural selection. Journal of Research in Science Teaching, 25, 733–746.

Linke, R.D. & Venz, M.I. (1978). Misconceptions in physical science among non-science background students. Research in Science Education, 9, 103–109.

Lucas, A.M. (1971). The teaching of "adaptation." Journal of Research in Science Teaching, 22, 261–278.

Malthus, T.R. (1798/1971). An essay on the principle of population, as it affects improvement of society. London: J. Johnson.

Mayr, E. (1982). The growth of biological thought: Diversity, evolution and inheritance. Cambridge, MA: Harvard University Press.

Mintzes, J.L., Wandersee, J.H., & Novak, J.D.(Eds.) (2000). Assessing science understanding: A human constructivist view (pp. 198–223). San Diego: Academic Press.

Morrison, J.A. & Lederman, N.G. (2000, April 28–May 1) Science teachers' diagnosis of students' perceptions. Presented at the annual meeting of the National Association for Research in Science Teaching, New Orleans, LA.

National Academy of Sciences. (1998). Teaching about evolution and the nature of science. Washington, DC: National Academy Press.

National Research Council. (1996). National science education standards. Washington, DC: National Academy Press.

Odom, A.L. & Barrow, L.H. (1995). The development and application of a two-tiered diagnostic test measuring college biology students' understanding of diffusion and osmosis following a course of instruction. Journal of Research in Science Teaching, 32, 45–61.

Otero, V. (2000). The process of learning about static electricity and the role of the computer simulator. Unpublished doctoral dissertation, University of California, San Diego, and San Diego State University.

Pedersen, S. & Halldén, O. (1994). Intuitive ideas and scientific explanations as parts of students developing understanding of biology: The case of evolution. European Journal of Psychology of Education, 9, 127–137.

Petren, K., Grant, B.R., & Grant, P.R. (1999). A phylogeny of Darwin's finches based on microsatellite DNA length variation. Proceedings of the Royal Society of London: Biological Sciences, 266, 321–329.

Renner, J.W., Brumby, M., & Shepherd, D.L. (1981). Why are there no dinosaurs in Oklahoma? Science Teacher, 12, 22–24.

Sadler, P.M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. Journal of Research in Science Teaching, 35, 265–296.

Scharmann, L. & Harris, W. (1992). Teaching evolution: Understanding and applying the nature of science. Journal of Research in Science Teaching, 29, 375–388.

Schluter, D. (2000). The ecology of adaptive radiation. Oxford, UK: Oxford University Press.

Settlage, J. & Odom, A.L. (1995, April 22–25). Natural selection conceptions assessment: Development of a two-tier test ''Understanding Biological Change.'' Presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco, CA.

Tabak, I. & Reiser, J. (1997, June). Domain-specific inquiry support: Permeating discussions with scientific conceptions. Paper presented at the symposium From Misconceptions to Constructed Understanding, Cornell University, Ithaca, NY.

Tamir, P. (1971). An alternative approach to the construction of multiple choice test items. Journal of Biological Education, 5, 305–307.

Taylor, W. (1953). Cloze procedure: A new tool for measuring readability. Journalism Quarterly, 30, 414–438.

Thorpe, R.S. & Brown, R.P. (1989). Microgeographic variation of the colour pattern of Canary Island lizards. Journal of the Linnean Society, 38, 303–322.

Treagust, D.F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. International Journal of Science Education, 10, 159–169.

von Glasersfeld, E. (1989). Cognition, construction of knowledge, and teaching. Synthese, 80, 121–140.

Weiner, J. (1995). The beak of the finch: A story of evolution in our time. New York: Vintage.