ARTICLE

Reasoning About Natural Selection: Diagnosing Contextual Competency Using the ACORNS Instrument

ROSS H. NEHM, ELIZABETH P. BEGGROW,
JOHN E. OPFER, MINSU HA

Abstract

Studies of students' thinking about natural selection have revealed that the scenarios in which students reason evoke different types, magnitudes, and arrangements of knowledge elements and misconceptions. Diagnostic tests are needed that probe students' thinking across a representative array of evolutionary contexts. The ACORNS is a diagnostic test that treats different evolutionary contexts as unique scenarios worthy of focused assessment and targeted instruction. Our investigations revealed that ACORNS scores produce valid and reliable inferences about students' thinking about natural selection. We urge biology teachers at all educational levels to begin assessing and attending to their students' reasoning across a broader array of evolutionary contexts, as competency in one context is often not indicative of competency in another.

Key Words: *Evolution; natural selection; contexts; diagnostic assessment; reasoning; learning.*

Evolutionary change is a central, observable feature of the natural world. For the past 150 years, the theory of natural selection has served as the primary (but not exclusive) explanation for evolutionary change (Endler, 1986; Gould, 2002). Specifically, natural selection is a mechanism that explains how the constant production of novel heritable variants – through the actions of mutation, genetic recombination, and sex – differentially persist from generation to generation through nonrandom survival and reproduction. Despite an expansive terminology and empirical body of work on natural selection, biologists agree that three core ideas are necessary and sufficient for explaining evolutionary change by natural selection: (1) the presence of variation; (2) the heritability of variation; and (3) the differential survival and/or reproduction of individuals that differ in heritable traits (Endler, 1986; Nehm & Schonfeld, 2010).

Evolutionary change is a central, observable feature of the natural world.

Despite its centrality in the life sciences, evolutionary change by natural selection is still poorly understood by students throughout the educational hierarchy (Gregory, 2009). This poor understanding has been attributed to a wide variety of cognitive, epistemological, religious, and emotional factors, yet there are still remarkably few tools available for validly assessing students' reasoning about natural selection (Nehm, 2006; Nehm & Schonfeld, 2008). This situation is problematic because quality assessments play a central role in helping teachers foster meaningful science learning (National Research Council, 2001), and they could play a similarly important role in improving students' understanding of how natural selection may be used to explain patterns of evolutionary change.

An important recent advance in assessment of natural selection has been the finding that the knowledge and misconceptions that students show vary greatly depending upon the specific contexts in which they are assessed (Nehm & Ha, 2011). For example, some students correctly explain the evolutionary *gain* of traits (such as the running speed of a cheetah) as being caused by the variability of the traits, their heritability, and the differential survival of organisms that possess the traits; however, these same students seldom mention these variables (variation, heritability, and differential survival) when explaining how traits *decline* in phenotypic frequency (such as the evolution of flightless birds). Indeed, understanding of one type of evolutionary change is a very poor predictor of understanding the other type. Despite these evident differences in students' own understanding of what is important in explaining evolutionary change, almost all existing assessments fail to probe students' thinking across the range of contexts in which evolutionary change actually occurs. Without assessing this range of contexts, how can teachers identify those instructional strategies that yield the broadest understanding of the chief cause of evolutionary change – natural selection?

Another problem with existing assessments is that they are inflexible, and their utility as diagnostic tools can degrade over time. As an example of this inflexibility, questions from widely used assessments – such as those by Bishop and Anderson (1990) and Anderson et al. (2002) – can become familiar to students after repeated exposure, and answers may even be disseminated among students. For teachers interested in understanding their students' reasoning about natural selection, we suggest that there are two fundamental problems that must be solved: (1) assessing contextual competence so that instruction can be planned accordingly and (2) having a tool that can be modified but retains inferences of validity. Here, we introduce a new diagnostic tool known as ACORNS (Assessing COntextual Reasoning about Natural Selection),

The American Biology Teacher, Vol. 74, No. 2, pages 92–98. ISSN 0002-7685, electronic ISSN 1938-4211.
DOI: 10.1525/abt.2012.74.2.6

provide evidence for its validity and reliability, and outline a methodology for teachers to modify the items and to use them as formative assessment tools in the classroom.

○ Natural Selection Reasoning Contexts

What are the contexts for reasoning about natural selection that we would like students to recognize? One might infer that they are – at a minimum – the ones addressed in curricula, and – ideally – the major contexts to which evolutionary reasoning applies. High school and college biology curricula seem to aim at a similar goal. They typically contain several different case studies of evolutionary change to illustrate aspects of natural selection and evolution. Well-known examples include Darwin's finches, bacterial resistance to antibiotics, and the evolution of horses (e.g., Campbell & Reese, 2008). Although a rationale for choosing the number and types of evolutionary case studies has never, to our knowledge, been explicitly justified or defended, a likely implicit rationale is that by exploring evolutionary change in a diversity of contexts students will progress toward an abstract conceptualization of natural selection that transcends particular cases. Such abstraction is also viewed as a central feature of knowledge *transfer* – that is, the ability to apply knowledge learned in one context to a different one (Barnett & Ceci, 2002; Opfer & Thompson, 2008).

How closely does students' reasoning meet the goal of an abstract conceptualization of natural selection? Recent work suggests that students fall short of this goal. Indeed, different biological contexts are associated with distinct patterns of student thinking, with student explanations very often depending on the superficial "cover stories" characteristic of evolutionary scenarios. Differences in reasoning may be revealed by comparing (Table 1): (1) within-species differences vs. between-species differences; (2) the gain of traits vs. the loss of traits; (3) familiar species vs. unfamiliar species; and (4) plants vs. animals (Nehm & Ha, 2011; J. E. Opfer et al., unpubl. paper). Typically, detecting students' knowledge and misconceptions in one context will *not* provide evidence of competency in another context.

These findings of context-dependent learning have a number of important implications for teaching evolution. First, curricula about evolution and natural selection require much care in the choice of the "cover stories" (such as bacterial resistance to antibiotics) that are used to illustrate evolutionary change. Ideally, such examples would represent a diversity of evolutionary scenarios that could be systematically compared and contrasted. In a variety of subject areas (e.g., mathematics), choosing examples that allow systematic comparisons is known to help students identify the variables that are truly important for problem solving (for a review, see Gentner & Colhoun, 2010), and we think it quite likely that the same would be true in learning the important variables that cause evolutionary change via natural selection. Additionally, "cover stories" might be chosen to reveal and address the naive ideas that plague student reasoning. Like students' understanding of the key variables in evolutionary change, misconceptions are also context-dependent, with misconceptions triggered by some contexts being rarely elicited by other contexts (Nehm & Ha, 2011).

No biologist would doubt that targeting meaningful learning of evolution across contexts is essential for making use of evolutionary theory. From this perspective, the crux of biology education is to foster effective evolutionary reasoning across *all* the branches on the tree of life, not just for a few disparate "twigs." In our view, the first step toward this goal is to employ diagnostic or formative assessment instruments that provide valid and reliable evidence about student thinking across diverse examples. Without such tools, it is simply impossible to know when an instructional intervention has provided students with the tools for making use of natural selection.

○ ACORNS

To better assess students' abilities to use natural selection to explain evolutionary change, we developed a new diagnostic instrument. The ACORNS is a short-answer diagnostic test modeled after Bishop and Anderson's (1990) widely used instrument. It builds on this prior work by explicitly delineating and expanding the contextual variables central to evolutionary reasoning (Table 1). It treats different "cover stories" as unique scenarios worthy of focused instructional attention. For example, questions prompt reasoning about the evolution of trait gains in familiar animals, unfamiliar animals, familiar plants, and unfamiliar plants because we know that students' understanding displayed in one scenario might lag his or her understanding in another (Opfer & Gelman, 2010).

Furthermore, unlike prior tests, the ACORNS standardizes taxon and trait familiarity among items so that these effects are not conflated with other factors. In novice learners, for example, reasoning about the dodder's haustoria is typically different from reasoning about penguin wings, whereas for experts it is not. Thus, multiple versions of the instrument may be assembled to examine particular reasoning patterns: between-species gain vs. loss (standardizing by animals of comparable familiarity); within-species differences for familiar vs. unfamiliar taxa/traits (standardizing by animals or plants); and so on. Such flexibility allows teachers to tailor the ACORNS to their own, unique curricula.

A final aspect of ACORNS is that it prompts students to formulate their explanations of evolutionary change from the standpoint of a biologist ("How would biologists explain…"). Some assessments are vague in regard to the vantage point from which a response is to be conceptualized, as well as the audience that the response is intended for. Students' informal explanations are likely to be quite different from their explanations employing academic discourse. Thus, in the ACORNS, students are explicitly prompted to reason and write using scientific language in their responses. In this way, we can separate students' scientific explanatory abilities from their personal beliefs.

While Bishop and Anderson's (1990) test and a modified version known as the ORI (Nehm & Reilly, 2007) have both been shown to produce valid inferences about evolutionary thinking, we also examined aspects of the validity and reliability of our new derivative items. Many different models might be used to establish inferences about the validity and reliability of diagnostic test scores (AERA, NCME & APA, 1999); we employed convergent testing to explore validity, and response consistency to examine reliability (Furr & Bacharach, 2008). Specifically, for convergent validity evidence, we compared 28 undergraduate students' (mean age 19.8; 60% female; 80% White non-Hispanic) performance on three different measures of evolutionary knowledge and misconceptions: (1) clinical interview scores derived from >10 hours of oral questioning (mean 19 minutes/student); (2) CINS multiple-choice test scores (Anderson et al., 2002); and (3) ACORNS short-answer test scores. The overall purpose of this work was to make sure that the ACORNS produced meaningful results that could be trusted by biology teachers.

○ Methods

The 20 multiple-choice items on the CINS were tallied as correct or incorrect (20-point maximum). Interview performance was scored on

Table 1. Differences in students' reasoning about natural selection.

Reasoning Contexts	Reasoning Patterns	ACORNS Items for Revealing Reasoning
Immunity/resistance vs. other trait changes	**Immunity** and **resistance** are unique reasoning contexts that elicit naive ideas about "adapting" similarly to adjustment or acclimation, whereas other trait-change scenarios do not elicit the same types of concepts or naive ideas.	**(A)** How would biologists explain how a living bed bug species with resistance to a pesticide evolved from an ancestral bed bug species that lacked resistance to the same pesticide? **(B)** How would biologists explain how a living mosquito species resistant to DDT evolved from an ancestral mosquito species that lacked resistance to DDT?
Within vs. between species differences	For the same taxon (e.g., birds), **within** a species, biological concepts such as mutation, sex, recombination, and heredity are commonly used by students to explain biological differences. By contrast, naive ideas are much more prevalent in **between-**species explanations (e.g., in birds).	**(A)** How would biologists explain how a species of flightless birds evolved from an ancestral bird species that could fly? **(B)** How would biologists explain how some individuals of flightless birds originated within a population of bird species that could fly?
Gains vs. losses of traits	For the same taxon/trait (e.g., rose thorns), students are typically much more adept at using scientific ideas to explain the **gain** of thorns, whereas significantly more naive ideas are invoked in situations involving the **loss** of, for example, rose thorns.	**(A)** How would biologists explain how a living rose species with thorns evolved from an ancestral rose species that lacked thorns? **(B)** How would biologists explain how a living rose species lacking thorns evolved from an ancestral rose species that had thorns?
Animals vs. plants	The types of naive ideas used to explain evolutionary change are typically different between **animals** and **plants.** Intentional and "use and disuse" explanations are more common for animal items, whereas teleological explanations are more common for plant items. Overall, plant evolution appears to be more difficult for students, perhaps because plants are often less familiar to students (see below).	**(A)** How would biologists explain how a living mouse species with claws evolved from an ancestral mouse species that lacked claws? **(B)** How would biologists explain how a living lily species without petals evolved from an ancestral lily species that had petals?
Familiar vs. unfamiliar taxa/traits	Students demonstrating competency in evolutionary reasoning using **familiar** taxa/traits often have difficulty abstracting their thinking to **unfamiliar** cases (e.g., dodder haustoria). Students often believe it is not possible to solve the problem without knowing how the trait functions, which likely indicates the absence of an abstract model of natural selection.	**(A)** Dodder, a plant species, have haustoria. How would biologists explain how the dodder species with haustoria evolved from the ancestral species that lack haustoria? **(B)** How would biologists explain how a living *Suricata* species that lacks a pollox evolved from an ancestral *Suricata* species that had a pollox?

a scale from –1 to +1, based on the overall magnitude of scientifically accurate or inaccurate responses (mirroring the methods of Nehm & Schonfeld, 2008). Interview questions included two ACORNS items, a CINS item, and two novel isomorphic items (see Appendices 1 and 2). ACORNS short-answer responses were tallied for the number of scientific key concepts (e.g., variation, heredity, differential survival) as well as naive ideas (needs, goals, use and disuse, etc.) using the scoring rubrics of Nehm et al. (2010). Reliabilities for the CINS and ACORNS were calculated using Cronbach's alpha (Ary et al., 2002).

○ Results

Our detailed analyses were used to determine how well the ACORNS exposed student thinking about patterns of evolutionary change by natural selection. Interview inter-rater reliabilities (using blind scoring) were 75%, and all scoring differences were resolved via deliberation between the raters. Kappa inter-rater reliabilities for ACORNS essay scoring were >0.80, discrepancies of which were also resolved via deliberation. No *scoring* reliability measures were needed for the CINS, as answers were either right or wrong relative to the answer key. Overall, different raters generated very similar assessment scores, which gives us confidence in scoring consistency.

Does ACORNS validly capture the thinking patterns of students? To answer this question, we compared performance on the ACORNS test to scores derived from an oral interview (considered the "gold standard" in education research) and the multiple-choice CINS test (Figure 1). The strong and statistically significant agreement between clinical interview scores and students' ACORNS scores supports validity inferences (Figure 1). Reliabilities, measured using Cronbach's alpha, were also robust and statistically significant (Key Concept Alpha = 0.77; Misconception

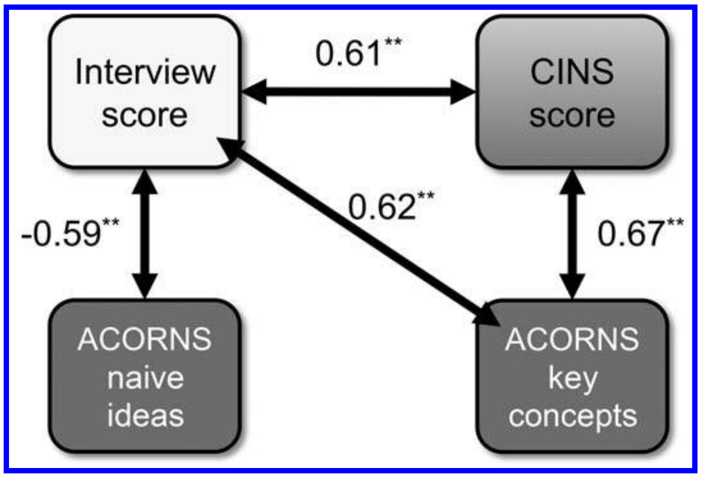

Figure 1. Convergent validity evidence for the ACORNS as measured by Pearson correlation coefficients among measured variables. The absence of a connecting line between variables indicates that there was not a significant association (**P< 0.01).

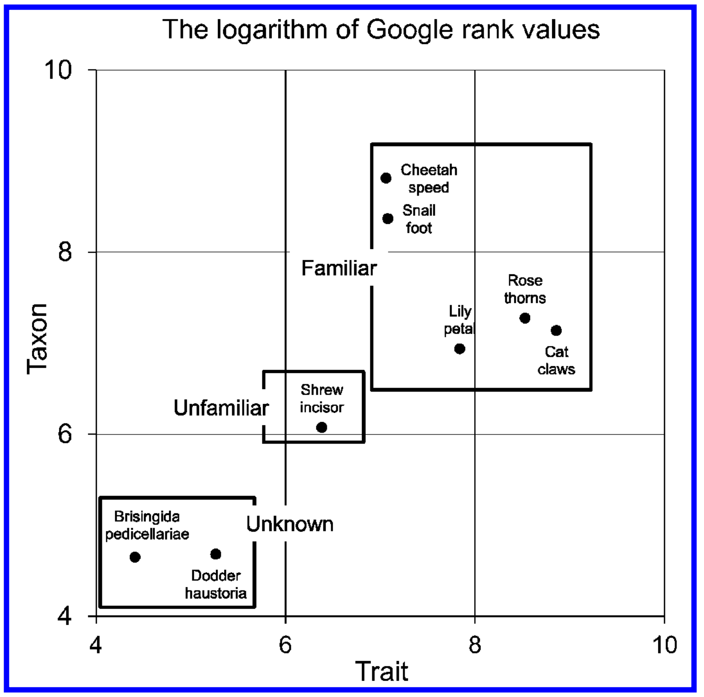

Figure 2. Estimates of taxon and trait familiarities using Google Ranks (listed in every Google search) on a logarithmic scale. Note that unfamiliar taxa/traits (e.g., dodder haustoria) are ranked much lower than familiar taxa/traits (e.g., rose thorns). Thus, Google Ranks and N-grams may serve as useful tools for assessing familiarity.

Alpha = 0.67; CINS Alpha = 0.75). Interestingly, although the number of naive ideas captured using the ACORNS was significantly and meaningfully associated with naive idea frequencies captured in clinical interviews with students, this was not found to be the case using CINS scores (Figure 1). Given that prior studies have also noted related problems with the CINS (Nehm & Schonfeld, 2008; Battisti et al., 2010), this result is not surprising. Overall, our results indicated that ACORNS scores served as valid and reliable proxies for students' reasoning abilities across different evolutionary contexts.

○ Implications for Biology Teachers

Our findings have a number of important implications for biology teachers that we highlight below.

(1) *Use ACORNS to assess understanding across multiple contexts*. The overarching implication of our work for biology teachers is that contexts or "cover stories" are significant factors in the teaching and learning of natural selection, and the ACORNS test may be used to expose students' context-specific reasoning patterns. Not all naive ideas and not all reasoning patterns will be exposed using one "cover story"; curriculum and instruction must be modified to address this fact. The ACORNS may be used as a diagnostic test before a unit on natural selection and thereby help align instruction with students' learning needs (National Research Council, 2001). Lessons about natural selection must not solely use examples of trait gains in familiar organisms (such as antibiotic resistance), but must also discuss cases of unfamiliar animals and plants, trait loss, etc. (Table 1).

(2) *Direct students' attention to how the same explanatory variables apply across the different items in ACORNS*. In teaching natural selection, it is helpful to include explicit comparisons across "cover stories" in order to help students identify the variables that are truly important for problem solving. By explicitly and systematically comparing the evolution of Darwin's finches to the evolution of antibiotic-resistant bacteria, it should be easier for students to see how the factors of variability, heritability, and differential survival operate to explain evolutionary change by natural selection. When insight is acquired across different examples that share few superficial features, it should help students use these same explanatory variables across different contexts.

(3) *Explicitly address student misconceptions*. Given that students' misconceptions (such as "use and disuse") often coexist with students' use of accurate key concepts (Nehm & Ha, 2011), teachers need to make a special effort to combat these misconceptions. Teaching that variation, heritability, and differential survival are *necessary* for evolutionary change isn't enough – these variables are also *sufficient* for explaining evolutionary change via natural selection. By contrast, the "need" for a trait (a common student explanation) is neither necessary nor sufficient for evolutionary change.

(4) *Using ACORNS as a base, develop a battery of worked examples for assessment and instruction*. For obvious reasons, test questions must be changed from time to time. The same is likely true of instructional materials, which should include both familiar and less familiar examples. The reason for this is that highly familiar explanations for repeatedly presented examples (e.g., peppered moths) are often memorized by students but not understood. To estimate how familiar students already are with particular species and traits, we recommend using GoogleLabs Books Ngram Viewer (GoogleLabs, 2011) or Google Ranks. In our study, the frequency of species and traits in Ngram, as well as ranks in Google searches, followed general patterns of taxon and trait familiarity that one would anticipate (Figure 2). This approach may be a useful starting point for building and attempting to standardize a new battery of taxon/trait combinations for instruction and assessment. Additional studies are being conducted to expand upon our compendium of taxon/trait combinations that are realistic and of comparable familiarity and difficulty (see Appendix 2).

One potential drawback of using the ACORNS is that it requires scoring students' written responses, which is more time consuming and requires more training and expertise in evolution than scoring a multiple-choice test (Nehm & Haertig, 2012). Nevertheless, a series of studies exploring automated computer scoring of written responses to ORI and ACORNS items have shown great promise (Nehm & Haertig, 2012; Nehm et al., 2012). Progress on these efforts can be followed online at http://evolutionassessment.org/.

References

AERA (American Educational Research Association), APA (American Psychological Association), & NCME (National Council on Measurement in Education). (1999). *Standards for Educational and Psychological Testing.* Washington, D.C.: AERA.

Anderson, D.L., Fisher, K.M. & Norman, G.J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, *39*, 952–978.

Ary, D., Jacobs, L.C. & Razavieh, A. (2002). *Introduction to Research in Education, 6th Ed.* Belmont, CA: Wadsworth-Thomson Learning.

Barnett, S.M. & Ceci, S.J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, *128*, 612–637.

Battisti, B.T., Hanegan, N., Sudweeks, R. & Cates, R. (2010). Using item response theory to conduct a distracter analysis on Conceptual Inventory of Natural Selection. *International Journal of Science and Mathematics Education*, *8*, 845–868.

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.

Campbell, N.A. & Reece, J.B. (2008). *Biology, 8th Ed.* San Francisco, CA: Pearson Benjamin Cummings.

Endler, J.A. (1986). *Natural Selection in the Wild.* Princeton, NJ: Princeton University Press.

Furr, R.M. & Bacharach,V.R. (2008). *Psychometrics: An Introduction.* Los Angeles, CA: Sage.

Gentner, D. & Colhoun, J. (2010). Analogical processes in human thinking and learning. In B. Glatzeder, V. Goel, & A. von Müller (Eds.), *On Thinking, vol. 2: Towards a Theory of Thinking* (pp. 35–48). Berlin: Springer-Verlag.

GoogleLabs (2011). Books Ngram viewer. Accessed online 21 January 2011 at http://ngrams.googlelabs.com/.

Gould, S.J. (2002). *The Structure of Evolutionary Theory*. Cambridge, MA: Harvard University Press.

Gregory, T.R. (2009). Understanding natural selection: essential concepts and common misconceptions. *Evolution Education and Outreach*, *2*, 156–175.

National Research Council. (2001). *Knowing What Students Know.* Washington, D.C.: National Academy Press.

Nehm, R.H. (2006). Faith-based evolution education? *BioScience*, *56*, 638–639.

Nehm, R.H. & Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching*, *48*, 237–256.

Nehm, R.H., Ha, M. & Mayfield, E. (2012). Transforming biology assessment with machine learning: automated scoring of written evolutionary explanations. *Journal of Science Education and Technology*, *21*, in press.

Nehm, R.H., Ha, M., Rector, M., Opfer, J., Perrin, L., Ridgway, J. & Mollohan, K. (2010). Scoring guide for the open response instrument (ORI) and evolutionary gain and loss test (ACORNS). Technical Report of National Science Foundation REESE Project 0909999. Accessed online 10 December 2010 at http://evolutionassessment.org/.

Nehm, R.H. & Haertig, H. (2012). Human vs. computer diagnosis of students' natural selection knowledge: testing the efficacy of text analytic software. *Journal of Science Education and Technology*, *21*, in press.

Nehm, R.H. & Reilly, L. (2007). Biology majors' knowledge and misconceptions of natural selection. *BioScience*, *57*, 263–272.

Nehm, R.H. & Schonfeld, I.S. (2008). Measuring knowledge of natural selection: a comparison of the CINS, an open-response instrument, and an oral interview. *Journal of Research in Science Teaching*, *45*, 1131–1160.

Nehm, R.H. & Schonfeld, I.S. (2010). The future of natural selection knowledge measurement: a reply to Anderson et al. *Journal of Research in Science Teaching*, *47*, 358–362.

Opfer, J.E. & Gelman, S.A. (2010). Development of the animate–inanimate distinction. In U. Goswami (Ed.), *The Wiley-Blackwell Handbook of Childhood Cognitive Development, 2nd Ed.* (pp. 213–238). Malden, MA: Blackwell.

Opfer, J.E. & Thompson, C.A. (2008). The trouble with transfer: insights from microgenetic changes in the representation of numerical magnitude. *Child Development*, 79, 790–806.

ROSS H. NEHM is Associate Professor in the School of Teaching and Learning and the Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, 1945 N. High St., Columbus, OH 43210; e-mail: nehm.1@osu.edu. ELIZABETH P. BEGGROW (beggrow.7@osu.edu) and MINSU HA (ha.101@osu.edu) are Ph.D. students in the School of Teaching and Learning, The Ohio State University. JOHN E. OPFER is Associate Professor of Psychology, The Ohio State University, 1835 Neil Avenue, Columbus, OH 43210; e-mail: opfer.7@osu.edu.

Appendix 1. Selected quotations from students representing each of the three possible interview scores and their corresponding CINS score.

Interview Score	Selected Quotations from Interview Responses	Selected Quotations from ACORNS Responses	CINS Score
1	"In this ancestral species of possum there would have to be a population, and within that population, the ones that had tails, some would either have a mutation that would give a smaller or almost no tail. And so that variant of opossum would exist in the population. And some sort of an environmental or even a sexual pressure would be placed on those possums that would favor the ones with shorter or no tails. So in the next generations, the possums that, I guess didn't have tails were more favored, so they had more offspring, so they were more fit, and the frequency of those genes would increase each generation until the population, I guess, didn't have tails at all."	"In a population of ancestral, tendril-less grapes some of the individuals had a tendril-like structure. This structure might have allowed those grapes to grow taller and have better support and were able to reproduce more. Because they were more fit to have more offspring, the trait for the tendril structure became more frequent in the population. After time and generations that trait was more and more frequent, and some individuals had even more effective tendril structure, which became more frequent. Eventually fully formed tendrils such as we see in the current population were present in all individuals of the new species."	20
0	"I think this question is kind of similar to the last question except the teeth part is easier to understand because that's something that's used with humans as well for food consumption. So a snail that didn't have teeth and now it has teeth, it's descendent has teeth…there might have been…does that mean that all of…when we say that like this is the ancestor does that mean that population is the ancestor or that single organism is the ancestor?…I don't know, I'm having a really hard time, I'm sorry…maybe the teeth in the ancestral species weren't used so they kind of like faded out in the population's genetics. But then something happened, and the few individuals that had teeth…I mean…I don't know…were the ones that evolved into this new species that had teeth."	"Tendrils help to anchor a plant to a branch or post and aid in it's horizontal and vertical growth. A plant with stem tendrils is less likely to be damaged by wind or displaced by animals. Tendrils could have increased the survival rate in grape species." "Generally petals serve as an attraction device for pollinators like birds and insects. A lilly without petals is probably in existence because it now utilizes wind to aid in pollination. It is possible that animal pollinators were not helping plant reproduce as well as the wind pollinated (no petal) type."	15
−1	"So I suppose the snail with teeth needed it maybe for eating purposes or, you know, as protection from predators. So overtime the snail that didn't have teeth needed to find something that could, you know, maybe it no longer had something that it can eat without teeth so it needed to evolve teeth in order to eat or like I said to fight off predators, so I guess in that way it would need teeth in the long run. So over time things change, so we went from, you know, not having legs to having legs because we needed them, you know as time progresses, years went on, or a large amount of time, the snail needed teeth as a means for survival."	"Biologists would explain this evolution by saying that grapes lacked tendrils and selective pressures in which it under went basically forced them to gain tendrils in order to survive." "The living snail species currently has teeth because it needed another mechanism to fight of predators possibly or a new mechanism for finding/eating food."	10

Appendix 2. ACORNS items and formats.

ACORNS Item Examples	Item Formats
Gain	How would biologists explain how a living (Taxon) species with (Trait) evolved from an ancestral (Taxon) species that lacked (Trait)?
Loss	How would biologists explain how a living (Taxon) species lacking (Trait) evolved from an ancestral (Taxon) species that had (Trait)?
Within species	How would biologists explain how some individuals of (Taxon) with (Trait) originated within a population of (Taxon) species that lacked (Trait)?
Between species	How would biologists explain how a species of (Taxon) with (Trait) evolved from an ancestral (Taxon) species that lacked (Trait)?
Familiar taxa + traits for use in each item	• Bacteria, antibiotic resistance • Cactus, spine • Cheetah, speed • Elm, winged seed • Fish, fins • Fly, wing • Grape, tendrils • Lily, petals • Locust, DDT resistance • Mouse, claws • Oak, nut • Opossum, tail • Penguin, flightless • Rose, thorns • Salamander, eyesight • Snail, foot • Snail, poison • Snail, teeth
Unfamiliar and unknown taxa + traits for use in each item	• Dodder, haustoria • Labiatae, Pulegone • Prosimian, tarsi • Shrew, incisors • *Suricata*, pollex