The of Logic of Adaptation:
How Do Students Reason About Natural Selection and Fit of Form and Function?

L. E. Delaney

April 19, 2020

"Seen in the light of evolution, biology is, perhaps, intellectually the most satisfying and inspiring science. Without that light it becomes a pile of sundry facts — some of them interesting or curious but making no meaningful picture as a whole." (Dobzhansky, 1973, pg. 129)

"Dobzhansky was right: let's tell the students." (Scott, 2010)

Introduction.

Evolution has long been regarded by biologists as the core principle undergirding all of life sciences. In the last few decades, substantial research efforts have attempted to identify the two-fold problem of evolution in biology education: why is evolutionary theory both difficult to understand and difficult to accept? Nearly one-third of adults in the United States still ascribe to Young Earth Creationism, and repeated studies have indicated that even after postsecondary education, fundamental misconceptions regarding the processes and patterns of evolution persist.

For thousands of years, humans have sought explanations for the origins of our natural world and the vast complexity of life it contains. Before the familiar One God, Seven Days, there were hundreds, perhaps thousands, of other stories confronting the difficult question of how organisms on Earth came to be. Many, many eggs — the infant god P'an Ku, floating perpetually in the inchoate sea of Chaos, the Egyptian sun god Ra, the Finnish diving duck, whose egg fragments on the knee of the goddess of air gave birth to the world. Creation *de novo*, *ex nihilo*, creation from chaos — what precipitated such, as Darwin so famously put it, "endless forms most beautiful and most wonderful" (1859, pg. 490)? What gave rise to these endless forms, and moreover, why are they so thoroughly suited to their environment?

Many of these perennial questions find answers within the processes of evolution. We now know that natural selection is the blind force that makes organisms seem as if they are "designed" expressly for their environment. It is the process that gives rise to adaptations: organisms appear engineered for specific tasks, interfacing with their surrounding ecology using remarkably honed sets of tools (Vincent & Brown, 2005). We refer to this concept as "fit of form and function" (FF&F): this idea — specifically, the observation that organisms are uniquely well-suited to their environments — was established long before the theory of evolution itself and is a common thread woven throughout religious, philosophical and scientific thinking for centuries. It is here that we aim to probe student understanding.

Understanding evolutionary theory is essential to building a robust conceptual framework in biology. And yet, there is little consistency in how and when evolutionary principles are taught: this can vary by state, by school, or even by individual instructor within schools. Some students are not exposed to evolutionary theory in their biology classrooms until after high school, while others begin learning these concepts in grade school.

Thus, students entering undergraduate introductory biology courses vary widely in their understanding of core biological principles. It is then the task of college instructors to educate a broad mix of students. This task is exceptionally difficult for both educators and their students given the lack of unifying framework in biological sciences generally, as there exists no codified outline for progression within the discipline. Biology majors progressing through coursework take classes in disparate departments (molecular biology, microbiology, ecology and evolution, etc.) that often do not highlight unifying themes and the interrelatedness of topics. Evolutionary principles are simply left for the course dedicated to evolution. Perhaps instead we should wonder: is it surprising that the general public does not view evolution as central to all of biology when it is almost never taught as such?

Students commonly complain that introductory biology courses require substantial memorization. This is unsurprisingly given the monumental job of introductory professors in preparing such a broad range of students with all the terms and concepts they may need in their future studies. However, recent work suggests that being taught a series of examples to memorize does not facilitate connections between facts — students rarely make these connections on their own, meaning they do not have the organizing framework necessary for subsequent study. Moreover, in a review of the three bestselling introductory biology textbooks, researchers found that the textbooks themselves serve to reinforce isolation of evolutionary concepts by presenting them solely in sections related specifically to evolution and rarely anywhere else (Nehm et al., 2009).

Educators have long been aware that students enter the classroom with a variety of ideas and beliefs about the natural world that are often at odds with established science. These "ideas and beliefs" have been the subject of a large body of research — also called "misconceptions", "alternative conceptions", "prescientific ideas", "intuitive beliefs" and more — that has repeatedly shown that they interfere with students' ability to appropriately understand and relate concepts. While thousands of studies have attempted to catalogue these misconceptions across many disciplines (see Duit, 2006 for a running bibliography of more than eight thousand papers), a growing body of work is attempting to understand the *how* and the *why*: what is it about the way we think and reason that causes these misconceptions to form, and why are they so resistant to formal training?

Natural selection is well-known as a concept in evolutionary theory that is particularly challenging for students. More than a dozen papers in the last thirty years (see Ziadie & Andrews, 2018 for review) have examined student thinking as it relates to natural selection. And yet, confusion regarding natural selection is tenacious, including among college graduates and teachers. Perhaps by focusing on the process of natural selection through the nuanced lens of FF&F, we can more carefully probe the underlying logic of students' beliefs. Until educators have a clearer picture of such logic, biology education will remain a difficult challenge. Here we propose an open-ended pre-post assessment to examine student thinking as it relates to FF&F in two populations of undergraduate students.

Background and motivation.

Evolutionary biology is widely regarded as the framework that organizes the entire discipline of biological sciences. To wit, there is no *why* in biology without underlying evolutionary processes. Despite this, evolution is often taught separately from other biological disciplines and sometimes not at all. Not only does this impede the construction of a conceptual framework in biological sciences generally, but the concepts required to fully understand evolutionary theory are also necessary for basic science literacy.

Fit of form and function.

Natural selection is a process that produces adaptations. Adaptations are traits that increase fitness for a particular organism, in a particular environment, relative to those that lack the trait. The logic behind this process is outwardly simple:

1. Individual organisms in a population vary — that is, they are not all exactly alike.
2. Some of these differences are heritable.
3. In each generation, organisms produce more offspring than are capable of surviving.

    Inference 1: There is a struggle for survival.

    Inference 2: Some organisms will survive and produce more offspring than other organisms.

4. Individuals with traits more conducive to survival and reproduction will appear in greater frequency in the subsequent generations. Thus, the surviving population represents a non-random sample of the original population.

Natural selection as described by Darwin in *Origin* (1859) is ultimately one (long) logical argument (Mayr, 1991). With no knowledge of genetics or understanding of the mechanisms of inheritance, Darwin was able to construct the argument behind his theory — he did so by focusing on the adaptations (traits) themselves and their distribution within environments (i.e., their fit of form and function). This suggests that understanding the nature of adaptations and how they may form via natural selection is an extremely powerful aspect of understanding evolution generally. In fact, many traits that appear superficially equivalent actually evolved separately, with completely different genes and physiology. What they have in common is not genes or underlying architecture, but selective pressure from similar ecological circumstances.

While genes are ultimately critical for understanding inheritance, they are clearly less critical understanding the process of natural selection. In fact, a strictly genetic approach to teaching natural selection will likely be insufficient for student understanding as this approach is almost completely devoid of ecology. Without the appropriate ecological context, the patterns produced by the process natural selection are difficult to interpret — why are organisms (and their associated adaptations) distributed as they are, and not some other way? Why are there so many organisms on Earth, and why are they all so different?

Understanding natural selection is essential to understanding biological diversity, species distribution, and trait distribution, but it is also essential for other disciplines vital to human survival like medicine and agriculture. Moreover, it can aid in building the kind of conceptual framework lacking for most undergraduate students in biology: instead of simply memorizing the differences between prokaryotic and eukaryotic cells, understanding their evolutionary relationship creates a foundation on which these facts can be constructed meaningfully.

All living organisms are open systems. That is, energy is constantly being exchanged between organisms and their immediate surroundings. Survival ultimately depends on reaching some kind of (dynamic) equilibrium with the environment. If an organism survives, gathers resources, and reproduces within such an environmental context, it suggests broadly that it is adapted to do so. Teasing out the selection pressures responsible for specific traits, determining what constitutes an "adaptation" versus a trait established by non-selective processes: these are difficult questions long grappled with in the science of evolutionary biology. Yet despite these thorny issues, many of the processes that shape organisms' FF&F are perhaps easier to tease apart. Indeed, many naturalists appreciated and studied FF&F long before the theory of evolution itself.

We suggest that the core of understanding adaptation and how adaptations form lies in what we will call the "6 Cs":

1. **Context-dependency.** Organisms inhabit an often changing physical environment where they must gather and use resources to survive and reproduce. Because living entities are open systems, they must also be capable of exchanging energy between themselves and their surroundings. The set of traits that allow them to achieve these goals (their "strategies") must incorporate both the context of these surroundings and the strategies of other organisms within and outside of their species. A drought-tolerant Mojave desert plant will not survive in the Swiss Alps.

2. **Constraints.** Genetic, developmental, and phylogenetic constraints determine the set of traits that selection may act on; so too do physical laws. Variation is the raw material of natural selection, meaning anything that limits the availability and structure of variation will constrain the possibilities available for adaptation. Studies suggest that the decrease of the oxygen in the atmosphere is at least partly responsible for the disappearance of large-bodied reptiles (Owerkowicz, Elsey, & Hicks, 2009).

3. **Compromises (i.e., trade-offs).** Once a particular strategy is selected in a given context by a particular group of organisms, other strategies are no longer available. Developmental and genetic constraints mean that organisms cannot be adapted to all contexts. A classic example are the compromises between survival, growth, and reproduction in life history theory.

4. **Co-adaptation.** Adaptations occur in the context of other adaptations within the organism itself. Traits do not exist within a vacuum; all traits must function together within an organism in a way that facilitates survival and reproduction. If not, that collection of traits will not persist. All organisms represent a collection of traits that are not necessarily the result of phylogenetic constraints.

5. **Co-evolution.** Adaptations can occur in response to the traits and strategies of other species. Two (or more) species can produce reciprocal selection pressures influencing each other's evolutionary pathways. It is suggested that the species richness of the orchid group may be a result of their stunning variety of highly specialized plant-pollinator relationships.

6. **Convergent evolution.** Organisms that do not share common ancestry may independently evolve similar traits in response to similar ecological circumstances. This can occur in two contexts. One is that an analogous trait or strategy is found between distantly related organisms, like echolocation in bats and whales. Another is that two distantly-related organisms may have traits that allow them to take advantage of the same opportunities. For example, speed is a common trait among organisms in grasslands, from kangaroos to pronghorns. As such, we may distinguish what is selected, and what is selected *for*: speed can be selected *for*, but what is *selected* — wings, flippers, paws — may be different.

A thorough understanding of these ideas is necessary for truly understanding FF&F, and would provide students with an appropriate framework to understand much of the process of natural selection and the patterns it can produce. Thus, FF&F provides a basis on which to probe how students think and reason as it relates to these concepts in evolutionary theory.

Previous work.

Despite its superficial simplicity, the process of natural selection and the resultant patterns of adaptation remain murky for students and educators for a variety of complex and interrelated reasons. A few lines of germane research examine the nature of misconceptions in science education from varying angles. diSessa (1993; 1998) suggests the existence of basic knowledge structures called phenomenological primitives (p-prims). These p-prims serve as a type of heuristic for reasoning through scientific questions — in the classic example, the fact that (1) heat comes from the sun and that (2) it is hotter closer to a heat source are both true beliefs that facilitate the (faulty) conclusion that it is hotter in the summer because the earth is closer to the sun. This is not a belief that students carry around in their day-to-day life (i.e., it is not part of a "faulty knowledge system" [Hammer, 1996, pg. 102]), but likely one they constructed in the moment based on simple facts previously derived through everyday experiences.

diSessa argues that such p-prims provide students with a collection of "recognizable phenomena" (1983, pg. 16) that allow them to develop explanations of their experiences — loosely connected p-prims related to a specific concept make up a "causal net" that helps students reason through a problem. However, other researchers suggest that misconceptions are not born out of causal nets but out of fully-realized intuitive conceptual frameworks (Coley & Tanner, 2015; Vosniadou, 1994). While these frameworks may provide the ability to understand, explain, and predict events in the natural world, they also often rely on faulty logic that can impede the construction of a more formal conceptual framework.

Much of the work examining p-prims relates to student understanding of physical concepts. What are the fundamental differences in how a naïve theory of physics is constructed versus a naïve theory of biology? The answer to this question is not clear. Some work with children suggests that the construction of these naïve theories and the modes of reasoning they employ are not the same (e.g., Coley & Muratore, 2012; Inagaki & Hatano, 2002). Yet, the possibility exists that as children and adolescents abstract and use p-prims in their daily life, they ultimately use the p-prims with the greatest predictive value to build more vigorous frameworks for understanding. These naïve frameworks would theoretically replace causal nets with a system that has a more organized underlying logic.

What are the ways students' logic may be flawed? Coley and Tanner (2015) address this question by providing students with six common biological misconceptions and asking them to provide justifications for their agreement or disagreement. The hypothesis is that specific modes of intuitive understanding and interpreting (conceptual frameworks called "cognitive construals") are responsible for the formation of many misconceptions specific to biology. Their results demonstrate frequent agreement with the biological misconceptions — among majors and non-majors alike — but more importantly, the associated justifications employed the specific construals hypothesized to generate such misconceptions.

Other lines of research examine the formation of misconceptions through the schemas they employ to understand processes. Work by Chi and colleagues (2012; 1998) suggests that students struggle with concepts relating to emergent processes like diffusion and natural selection. Their hypothesis is that early reasoning skills often utilize a Direct-causal schema, or rather, a "generalized version of narrative schemas and scripts" (Chi et al., 2012, pg. 1). This borrows from two lines of evidence. One is that children comprehend stories through a simple schema that contains components necessary for understanding basic narratives: *triggering event*, *protagonist*, *series of actions*, ultimately leading to a *conclusion*. The other relates to how both children and adults understand familiar everyday events through the construction of a "script" (Schank & Abelson, 1977). These scripts facilitate both understanding and planning action by allowing an individual to reason through the few events that may occur in any given situation. For instance, a "grocery store script" would allow an individual to reason through the plans and ultimate goal of obtaining food based on previous experiences at the grocery store (collecting items, bringing them to the register, paying, and so on).

Emergent processes are particularly difficult for students to grasp. Unlike a sequential process, where agents act in consecutive and distinctive ways (e.g., a baseball game), emergent processes include the action of agents acting simultaneously, independently, *and randomly*, from which a pattern "emerges". Using a Direct Schema (appropriate for sequential and stage-like biological processes like the menstrual cycle or DNA replication) leads to robust misconceptions when applied to an emergent process. It precludes viewing the resultant pattern as the collective summing of all interactions between agents at each point in time (Chi et al., 2012). Without an appropriate Emergent Schema to apply to these processes, students' logic relating to the process of natural selection and resultant patterns of FF&F will be flawed. Emergent processes may be especially difficult for students to grasp because (1) a Direct Schema is not sufficient for understanding emergent processes and their resulting patterns, but also (2) the great utility of causal nets or cognitive construals does not always extend to formal biological thinking, which is often counterintuitive. Despite formal instruction, students still have trouble understanding how a non-directed process may lead to the stunning varieties of FF&F.

Kalman, Morris, Cottin, and Gordon (1999) argue that students can be successful in their courses throughout the majority of high school simply by "memorizing templates for every situation encountered on an examination" (p. S45). That is, students are accessing templates constructed of rote facts dependent on the specific circumstances of the question, but not building a unified framework characteristic of an organized knowledge system. This finding supports anecdotal evidence from educators that changing the superficial details of a problem confuses students to the extent that they can no longer solve the problem.

In physics, Hammer (1989; 1994) has described how students view the subject as disparate pieces of information that are weakly connected — in opposition to the interrelated concepts their professors believe they are teaching. This lack of uniformity is likely exacerbated in biological sciences as the discipline itself is far from unified. Hestenes (1998) writes that while approximately 80% of typical university physics students could *state* Newton's Third Law prior to instruction, less than 15% fully understood the *concept* by the end of the course. This discrepancy was uncovered after administering the Force Concept Inventory (FCI) to hundreds of university and high school students before and after physics instruction.

The FCI (Hestenes, Wells, & Swackhamer, 1992) was based on an instrument that probed not intelligence, but student conceptual understanding of Newtonian physics. Shocking their physics professors, students generally performed quite poorly on the test despite the formal instruction they received. The explosion of concept inventories across disciplines in subsequent years indicates the popularity of these tools. The general purpose of a concept inventory is to test conceptual understanding in key areas through a multiple-choice test containing known "distractors" that are based on common misconceptions (Sands, Parker, Hedgeland, Jordan, & Galloway, 2018). In biology, concept inventories have been developed on subjects as narrow as genetic drift (GeDi), natural selection (CINS), or meiosis (Meiosis CI), and as wide as middle-school life science (MS-LSCI).

Concept inventories are generally an appropriate tool to examine the efficacy of particular teaching strategies. However, while these multiple-choice assessments may be valuable in diagnosing common misconceptions to specific problems, they fall short in assessing *why* a student may believe a particular idea or *how* they are employing this logic to other problems of a similar type. Moreover, students may memorize enough information to recognize distractor choices deemed "incorrect" by their teachers, but still not have a clear understanding of *why* those answers are incorrect. Probing student understanding thus requires space for students to organize their thinking and construct their own argument.

Uncovering student logic.

All of these lines of research suggest that while many factors are responsible for the lack of understanding and acceptance of evolution generally, there are a few common themes that likely have the most impact. Moreover, cognitive construals, along with inappropriate schemas and on-the-fly knowledge construction using p-prims ("causal nets"), likely all interact with each other in ways that are difficult to tease apart. This stems in part from the broad nature of the theory of evolution itself. In fact, the theory of evolution actually contains five theories as Mayr (1982) points out, further adding to its complexity. Thus, by examining a far narrower window FF&F, we may bypass many of these documented and difficult challenges and their potential interactions.

Writing is a process that reinforces reflection, connection, and organization (Durst & Newell, 1989). Proponents of the write-to-learn (WTL) movement argue that reflective writing enhances metacognition, an important aspect of constructing knowledge in a cohesive framework. Not only is reflective writing a useful tool for student learning and self-evaluation, but reflective writing can also be a useful tool for educators to diagnose misconceptions and the reasoning behind those misconceptions. Thus, posing a series of open-ended questions may allow students the space to elaborate their thoughts more clearly, giving a greater degree of insight into the modes of reasoning they employ, in addition to the factual information behind their logic.

More recently, methods of assessing student understanding that use open-ended questions have been published, like the Assessment of Contextual Reasoning about Natural Selection (ACORNS) instrument (Nehm, Beggrow, Opfer, & Ha, 2012). ACORNS is unique for a few reasons. First, it is one of only two (out of eighteen) evolution-specific concept inventories that employ open-ended questions. Second, the computational tool EvoGrader (Moharreri, Ha, & Nehm, 2014) was developed to automate grading of ACORNS and provide detailed feedback on student responses. This is a tremendous advantage for educators responsible for large classes, as evaluating written feedback to exams can be extremely time-consuming.

The ACORNS instrument is novel in respect to its use of open-ended questions. It is also an extremely useful tool for educators responsible for assessing large classes. However, it is somewhat limited in its ability to uncover student logic, as the software contains previously established "categories" of normative and non-normative ideas in which student responses are binned. It is not meant to be an exploratory tool, but one that simply assesses the extent to which student thinking conforms to the dominant conceptions in the field, and where students hold common and previously established misconceptions.

Open-ended response instruments are not without their own limitations. In addition to time spent grading, student responses may also be difficult to interpret for a variety of reasons (Nehm & Schonfeld, 2008). Some of the limitations relate to how seriously the students take the assessment and how much they elaborate on their responses. Short answers can be difficult to analyze in meaningful ways. Other limitations relate to the science of evolution itself, and the lack of consensus — even among evolutionary biologists — about the definitions of certain terms or the interpretation of certain facts. This means that the grading of these assessments may change depending on the individual grader. Regardless of these limitations, open-ended response assessments can still facilitate a detailed picture of student thinking that multiple-choice assessments are unable to provide.

Study populations.

Undergraduate Introductory Biology is unique among introductory science courses in that many non-science majors enroll to earn mandatory science credits. As such, these classes usually contain a heterogenous mix of students — non-STEM majors, STEM majors, and biological science majors that likely have varying degrees of formal biological education. In contrast, upper-level classes like Ecology tend to consist of more homogenous populations of upper-division students planning to pursue biological careers.

At University of Illinois at Chicago (UIC), Biological Sciences majors are required to take one course throughout their study that focuses on evolution (Ecology and Evolution, BIOS 230), generally taken sophomore or junior year. Nearly all students come from the state of Illinois, the majority of which come from the surrounding Cook County (UIC, 2020). High schools throughout Cook County vary widely, ranging in SAT school composite rank from 1375 to 741.1 (Tribune, 2017). Thus, the study populations will vary widely in their scholastic background and their exposure to biology and biological principles prior to postsecondary education. This study provides a unique opportunity to see how these differences may affect their understanding and learning outcomes in an introductory setting and several semesters into their degree.

In an ideal world, students in upper-division biology classes would exhibit a much more sophisticated understanding of biological processes like natural selection, and their resultant patterns like adaptation. However, many studies indicate this is not the case, demonstrating that even established professors can hold tenacious misconceptions relating to natural selection. Evaluating student thinking in these different populations is important because it allows for comparison across varying levels of formal biological education. It is essential for college educators to first understand the nature of these misconceptions and their underlying logic so that they may devise successful ways of addressing them in their classrooms.

Research overview.

Aims and broad overview.

Aim 1: How familiar are students with the concept of "adaptations" prior to postsecondary instruction? Does introductory coursework improve understanding of this and related concepts? Are there consistent misconceptions across question types, and do they utilize similar logic? How does understanding adaptation and related concepts relate to student learning outcomes (SLOs) and retention?

1. Provide BIOS 120 students with a series of open-ended questions regarding adaptations and the action of natural selection. These should incorporate common misconceptions that may stem from difficulty with any one or combination of the six Cs.
2. Complete this evaluation at the beginning and end of the course — track by individual.
3. On the individual level: compare pre- and post-test with course outcomes like exam grades, course grades, retention.
4. On the course level: evaluate potential changes between pre- and post-tests; evaluate pre- and post-tests (and changes) with SLOs and retention; determine difficult concepts and common conceptual errors.

Aim 2: How familiar are students with the concept of "adaptations" after 2-3 years of postsecondary instruction? After progression through their majors, do students show improved understanding of this and related concepts? Which misconceptions are the most persistent, and do they utilize similar logic? How does understanding these concepts relate to SLOs and retention?

1. Provide BIOS 331 students with a series of ten open-ended questions regarding adaptations and the action of natural selection. These should incorporate common misconceptions that may stem from difficulty with any one or combination of the six Cs.
2. Complete this evaluation at the beginning and end of the course — track by individual.
3. On the individual level: compare pre- and post-test with course outcomes like exam grades, course grades, retention.
4. On the course level: evaluate potential changes between pre- and post-tests; evaluate pre- and post-tests (and changes) with SLOs and retention; determine difficult concepts and common mistakes.

Methods.

Test validation.

The goal of the pre- and post-test is to examine the ways students reason about natural selection and FF&F — do their answers exhibit common misconceptions? What is the logic behind their misconceptions? Do the same misconceptions employ similar modes of reasoning? Does their score on this assessment relate to their overall course performance? There are currently five concept inventories that focus on the process of natural selection — three multiple-choice and two open-ended response assessments. Use of such assessments that look at evolution generally have generated insights on the breadth of misconceptions held by students and the modes of reasoning that underlie them. Several general patterns emerge from this literature: many misconceptions stem from teleological (goal-driven), essentialist, and anthropocentric modes of thinking (Coley and Muratore 2012;

2015); students may lack understanding of "threshold concepts" including probability, randomness, stochasticity, and timescale (Göransson, Orraryd, Fiedler, & Tibell, 2020); Lamarckian mechanisms of inheritance, species-level reasoning, and a variety of difficulties in explaining trait loss and gain (especially in unfamiliar organisms) also generate misconceptions (Nehm, 2018).

While these certainly play a role in student understanding of natural selection, focusing on the products of natural selection (the adaptations themselves) may allow students to understand the process while avoiding some of these common pitfalls. We propose that the six Cs — context-dependency, compromise, constraints, co-adaptation, co-evolution, and convergent evolution — are necessary and sufficient to provide students with a robust understanding of adaptation and the process of natural selection. Thus, the assessment questions, while open-ended, will focus on this specific and narrow window of evolutionary processes, specifically within the context of the six Cs.

Because the assessment focuses more on student logic than whether they provide the "right" answer, the questions will be broad and open-ended, leaving plenty of room for student interpretation. While the assessment will contain no more than five questions, at least 50 potential questions will be developed for review by content experts. Content experts include professional evolutionary biologists and seasoned educators in biological sciences. Ideally content experts will agree on at least ten open-ended assessment questions that can be used for the pre- and post-test in the Fall semester. Two versions will be developed and both versions will be administered in BIOS 120 and BIOS 331. After data analysis in the Fall semester, questions that garner responses with limited value will be removed or edited before the assessment is administered again in the spring.

Test administration.

Biology of Populations and Communities (BIOS 120) is an introductory survey course that covers broadly the following topics: species concepts, systematics, macroevolutionary processes such as extinction, transmission genetics, the origins and maintenance of genetic variation, natural selection and adaptation, other evolutionary processes including genetic drift, population biology and community ecology, worldwide biogeography, biomes, and biodiversity, biogeochemical cycles, conservation biology, the scientific method, and human issues associated with extinction. The course includes two weekly lectures and one weekly laboratory section. The class generally contains between 400-500 students per semester split into at least ten laboratory sections. Students are usually freshman or sophomores and a mix of biological sciences majors and non-majors.

General Ecology Laboratory (BIOS 331) is an upper-division ecology course that focuses on empirical methods with hands-on activities and field trips, covering: forest ecology, species interactions, life history theory, community ecology, conservation biology, succession, nutrient cycling, climate change, evolutionary ecology, and human ecology. It includes two weekly lectures and one weekly laboratory, in addition to three all-day field trips. The class usually contains approximately 100 students split into four laboratory sections. This is an upper-division class with students that are usually juniors or seniors, the majority of whom are biological sciences majors.

The assessment will be administered as a fifteen minute open-ended paper questionnaire given in the first and last laboratory (all sections) of the semester in both BIOS 120 and BIOS 331. Teaching assistants (TAs) will be instructed to give the assessment similarly to a quiz — they will not provide help to the students as they work their way through the questions. Students will receive an identical pre- and post-test.

Additional data.

In a study examining predictors of success in "gateway courses" (high rates of DFW grades), Benford and Gess-Newsome (2006) found three significant predictors for all science and mathematics gateway courses, and one predictor specific to biology. For all science and mathematics courses, quantitative and analytical skills (measured by SAT or ACT score), academic self-esteem (determined qualitatively via survey), and GPA were found to be significant predictors. In biology specifically, prior experience with course material and robust verbal skills were shown to also be strong predictors of success. Other studies have also shown that performance in high school science predicts future achievement in college coursework (Szabo, 1969).

Thus, in addition to the open-ended questions, students will be asked to provide additional data on their educational background. Students in BIOS 120 may provide data on their high school and what major they have declared, if any. BIOS 331 students may provide data on biological sciences classes they have taken previously and what major they have declared.

This additional data is important for two reasons. First, it will allow for binning of students into groups of majors and non-majors, study populations that should theoretically display different levels of complexity in their biological understanding. Second, it will give an indication of how much an individual student has been exposed to biological concepts and associated reasoning skills prior to enrolling in the class. As students come to UIC from schools of varying scholastic rank, gathering data on their background will likely be an important aspect of their biological understanding and learning outcomes in biological sciences.

Coding student responses.

As the questions are as open-ended as possible, student responses will likely vary substantially. Thus, these responses will be coded in two ways. The first will address the overall complexity of their responses using Webb's Depth of Knowledge (Webb, 2002, DoK). A level one response will differ from that of a level three response by extending from factual recall to synthesis and justification. Table 1 below outlines the four levels of depth and how they may be used to categorize student responses.

The second way to code student responses will address the modes of reasoning that potentially lead to faulty conclusions, i.e., the nature of their flawed logic (Table 1). Outlined originally by Nehm, Rector, and Ha (2010), coding students' cognitive reasoning models was adapted for the automated grading of ACORNS responses with the program EvoGrader (Moharreri et al., 2014). This broad binning will be useful for the variety of responses that will likely be collected, while maintaining the possibility for further characterization by specific naïve belief (e.g., which of the six Cs underly their flawed logic).

Students will receive a composite score based on DoK level and reasoning model. The higher the DoK level and the more "scientific" the reasoning model, the higher the composite score. For example, a complex answer that involves higher-order reasoning and (normative) ideas from multiple content areas *and* that employs a scientific mode of reasoning would receive the highest score of seven. An answer that simply restates the question without any underlying reasoning would receive the lowest score of one (Table 3).

Comparing student scores to student learning outcomes and retention.

Once the pre- and post-test have been administered, data on student learning outcomes will be collected for comparative purposes. Student learning outcomes are limited to retention, exam grades, and overall course grade. These metrics will be evaluated against student performance (score) on the pre-post test, in addition to their *change* in score from the beginning to the end of the class, if any.

The pre-post test scores and the change in score will be evaluated by major and non-major, high school rank or number of biology courses taken previously, exam grades, and final grades. Some students will receive a W or no grade (if they dropped the course early), so these pre-test scores will be evaluated against the individuals that remained in the course (Table 4).

Level	*Skills*	*Description*
1	Recall & Reproduction	Recall facts, definitions, principles, or concepts; display the least complexity in response; single-step answers or answers with circular logic (e.g., repeating the question in a new way).
2	Skills & Concepts	Use information and conceptual knowledge beyond simple recall; select appropriate justification for a claim; more complex reasoning (e.g., multi-step response).
3	Strategic Thinking	Reason or develop a plan to approach a problem; employ decision-making and justification; solve abstract, complex, or non-routine problems; identify multiple correct answers for a problem.
4	Extended Thinking	High cognitive demand and complex reasoning to make multiple connections; students relate ideas within and/or among content areas; indicate how an answer may be determined empirically.

Table 1: Webb's Depth of Knowledge (DoK), adapted from K. Hess, 2010; K. K. Hess et al., 2009

Modes of Reasoning	*Definition*	*Examples*
Pure scientific	Logic adheres to the dominant conceptions within the field.	Answers contain only normative scientific ideas and reasoning.
Mixed	A mixture of both normative scientific reasoning and naïve reasoning	Any mixture of naïve reasoning and scientific reasoning or ideas
Pure naïve	Logic does not adhere to dominant conceptions in the field but does adhere to some kind of intuitive reasoning.	Misunderstanding of the six Cs; essentialist, teleological or anthropocentric thinking; direct-causal schema; causal nets.
None	No reasoning model employed.	Re-stating the question, providing extraneous information, not answering the question directly.

Table 2: Adapted from Moharreri et al., 2014; cognitive reasoning models taken from Nehm et al., 2012.

The goal of this comparison is not to assess the "success" of the classes themselves at teaching natural selection. The classes are chosen not necessarily for content, but because of the populations of students they contain — a mix of majors and non-majors in BIOS120, and upper-division majors in BIOS331. We expect that BIOS120 students may exhibit a greater change between the pre- and post-test because this course deals with many of these concepts more specifically than does BIOS331. However, students that display a more sophisticated understanding of these concepts in their pre-post test are perhaps likely to have better learning outcomes. Longitudinal studies of this kind are rare and more data is needed to assess that hypothesis.

Score	*Models + DoK level*	*Examples*
1	No reasoning model employed (0) + DoK level 1	Re-stating the question; answering a different question than the one asked; no answer.
2	Pure naïve model (1) + DoK level 1	Reproduction or recall of some non-normative ideas with intuitive biological reasoning.
3	Mixed reasoning model (2) + DoK level 1; pure naïve model (1) + DoK level 2	Recall of facts with any mixture of naïve reasoning and scientific reasoning; intuitive biological reasoning with some justification or synthesis beyond recall.
4	Pure scientific model (3) + DoK level 1; mixed reasoning model (2) + DoK level 2	Recall or reproduction of scientific facts and principles; some scientific and naïve reasoning models with answers that display some kind of synthesis of information.
5	Pure scientific model (3) + DoK level 2; mixed reasoning model (2) + DoK level 3	Synthesis of scientific facts and principles; Some mixture of scientific and naïve reasoning with deeper evaluation, including identifying multiple correct answers.
6	Pure scientific model (3) + DoK level 3; mixed reasoning model (2) + DoK level 4	Synthesis, justification and interpretation of scientific facts and principles, possibly including more than one correct answer; highly complex response with multiple connections across content areas, mixed justification of claims, suggestions of empirical work to determine answers.
7	Pure scientific model (3) + DoK level 4	Highly complex response with multiple connections across content areas, appropriate justification of claims, suggestions of empirical work to determine answers.

Table 3: Examples of levels of coded student responses.

Possible end SLOs	**Category**
Grade A	Pass – High
Grade B	Pass
Grade C	Pass – Low
Grade D or F	Fail
Grade W	Withdraw
No grade	Withdraw

(a)

Additional data	**Category**
Biological sciences major	Major
Other major	Non-major
High school (categorical)	Previous exposure
College course completion (number)	Previous exposure

(b)

Table 4: Additional data gathered on (a) final student learning outcomes and (b) student demographics.

Comparing student scores between BIOS 120 and BIOS 331.

How did students' pre-post test grades differ between courses? Did BIOS 331 students employ more scientific reasoning modes than comparable students in BIOS 120? To address these questions, we will compare pre-post test grades among students of similar status that achieved similar learning outcomes (grades, major/non-major, etc.). This will give an indication of how student thinking may change after several semesters of postsecondary study. We may also expect that students in BIOS 120 will show greater gains between the pre- and post-test as many of them will be exposed to biological principles that they may not have seen previously. This differs from students in BIOS 331 that have already taken several classes in biology.

Modify and repeat procedure in spring semester.

Once student responses have been gathered and analyzed, the pre-post test can be modified for the following spring. Some questions will likely have limited utility for a variety of reasons. Questions that gather limited usable data (e.g., the majority of students did not attempt them) will be removed. Questions that garner confused responses from the majority of students may also be edited for clarity. Once these changes are made, the new version of the test will be administered again in the same classes with the same procedure during the spring semester.

Implications.

This study will aid in improving biology education in a four ways. First, only a few studies to date have examined the underlying logic students use to answer questions about natural selection, and none have focused specifically on adaptations themselves. This study will be the first of its kind to examine this through open-ended pre-post tests. Second, there is a great need for longitudinal data in education research. Many studies that involve pre-post tests simply eliminate the students that dropped the course without including those statistics. Here we propose to track those students along with their pre-test score, giving us important data on the retention rate in these classes. Third, few studies have looked at differences between majors and non-majors, and it would be helpful to see if previously discovered patterns hold across institutions. Fourth, and perhaps most importantly — are we teaching biology students how to think about biology? While this question cannot be answered by any one empirical study, these kind of assessments can aid educators in forming a better picture of how our students think and reason (or don't) about biological concepts.

Appendix.

Running list of possible questions for assessment.

True false questions — explain in a few sentences.

1. A group of fish within the same school are essentially the same, except for their age differences.
2. Only the very best organisms in a group of species will survive to reproduce.
3. Over time, natural selection works to create a balanced ecosystem, where each species is optimized for its particular role.
4. If the environment stays exactly the same, natural selection cannot produce adaptations.
5. As their environment became snowy, rabbits developed new genes that gave them white fur.
6. Mammals are all adapted to have four limbs.
7. Without evolutionary change, a species will go extinct.
8. Organisms cannot adapt to climate change because natural selection works very slowly.
9. Because humans no longer face challenges to survival and reproduction, natural selection no longer acts in human populations.
10. Asexual organisms cannot evolve.
11. All organisms' traits are designed for a purpose.
12. Natural selection acts gradually, so it is impossible to see its effects over a human lifetime.

Open-ended questions.

13. Why is there wood?
14. How would biologists explain how a species of cactus without spines evolved from a species of cactus with spines? (Nehm, 2018)
15. If penguins are unable to fly, why do they have wings? Are these "wings" an adaptation?
16. Some species are found in highly localized areas (endemic), while others are found across the globe (cosmopolitan). For example, *Dubautia* is a genus of flowering plants found only in the Hawaiian islands, while plants in the genus *Ranunculus* are found on every continent on earth. What might explain these differences in distribution?
17. How do predators avoid eating all of their prey?
18. Do all organisms have a carrying capacity?
19. How do beneficial traits arise in a population of organisms?
20. If fish live in a dark cave for several generations they eventually lose their eyesight. Describe how this may happen over time.
21. At what point does one species become two species?

22. Is there a limit to how tall a tree can be?
23. Why do birds of the same species have different color coats?
24. Why are there so many different kinds of flowers?
25. Which of these species share a common ancestor: seagull, brown bat, whale, salmon, honey bee, tulip?
26. Why are there generally an equal number of males and females in a breeding population?
27. How is the process of breeding dogs with different characteristics different than the process of natural selection?
28. If biologists wanted to speed up evolutionary change, how would they do it? (Nehm & Riley, 2007)
29. Assuming the common ancestor of cheetahs ran no more than 20mph, explain in a few sentences how cheetahs are now able to run 60mph.
30. The male peacock's tail is extremely heavy and makes it difficult to fly except for short distances. Is this tail an adaptation? Why or why not?
31. If natural selection brings species closer to an optimum, then why are members of the same species not exact copies of each other?
32. Instead of having one organism perfectly adapted to life on earth, there are millions. Why is this the case?

References

Benford, R., & Gess-Newsome, J. (2006). Factors affecting student academic success in gateway courses at Northern Arizona University. *ERIC*.

Chi, M. T., Roscoe, R. D., Slotta, J. D., Roy, M., & Chase, C. C. (2012). Misconceived causal explanations for emergent processes. *Cognitive Science*, *36*(1), 1–61.

Coley, J. D., & Muratore, T. M. (2012). Trees, fish, and other fictions. In *Evolution Challenges: Integrating Research and Practice in Teaching and Learning about Evolution* (pp. 22–46). New York, NY: Oxford University Press.

Coley, J. D., & Tanner, K. (2015). Relations between intuitive biological thinking and biological misconceptions in biology majors and nonmajors. *CBE—Life Sciences Education*, *14*(1), ar8.

Darwin, C. (1859). *On the Origin of Species: A Fascimile of the First Edition.* Harvard University Press. (Copyright 1964.)

diSessa, A. A. (1983). Phenomenology and the evolution of intuition. In D. Gentner & A. Stevens (Eds.), *Mental Models* (pp. 15–33). Mahwah, NJ: Erlbaum.

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction*, *10*(2-3), 105–225.

diSessa, A. A., & Sherin, B. L. (1998). What changes in conceptual change? *International Journal of Science Education*, *20*(10), 1155–1191.

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, *35*(3), 125–129.

Duit, R. (2006). *Students' and teachers' conceptions and science education.* Retrieved from `www.ipn.uni-kiel.de/aktuell/stcse/stcse.html`

Durst, R. K., & Newell, G. E. (1989). The uses of function: James Britton's category system and research on writing. *Review of Educational Research*, *59*(4), 374–393.

Ferrari, M., & Chi, M. T. (1998). The nature of naïve explanations of natural selection. *International Journal of Science Education*, *20*(10), 1231–1256.

Göransson, A., Orraryd, D., Fiedler, D., & Tibell, L. A. (2020). Conceptual characterization of threshold concepts in student explanations of evolution by natural selection and effects of item context. *CBE—Life Sciences Education*, *19*(1), ar1.

Hammer, D. (1989). Two approaches to learning physics. *The Physics Teacher*, *27*(9), 664–670.

Hammer, D. (1994). Epistemological beliefs in introductory physics. *Cognition and Instruction*, *12*(2), 151–183.

Hammer, D. (1996). Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions. *The Journal of the Learning Sciences*, *5*(2), 97–127.

Hess, K. (2010). Applying Webb's depth-of-knowledge (DoK) levels in science. *National Center for Assessment*, *10*.

Hess, K. K., Jones, B. S., Carlock, D., & Walkup, J. R. (2009). Cognitive rigor: Blending the strengths of Bloom's Taxonomy and Webb's Depth of Knowledge to enhance classroom-level processes. *ERIC*.

Hestenes, D. (1998). Who needs physics education research!? *American Journal of Physics*, *66*, 465–467.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, *30*(3), 141–158.

Inagaki, K., & Hatano, G. (2002). *Young Children's Naive Thinking about the Biological World.* Psychology Press.

Kalman, C. S., Morris, S., Cottin, C., & Gordon, R. (1999). Promoting conceptual change using collaborative groups in quantitative gateway courses. *American Journal of Physics*, *67*(S1), S45–S51.

Mayr, E. (1982). *The Growth of Biological Thought: Diversity, Evolution, and Inheritance.* Harvard University Press.

Mayr, E. (1991). *One Long Argument: Charles Darwin and the Genesis of Modern Evolutionary Thought* (Second ed.). Harvard University Press.

Moharreri, K., Ha, M., & Nehm, R. H. (2014). EvoGrader: An online formative assessment tool for automatically evaluating written evolutionary explanations. *Evolution: Education and Outreach*, *7*(1), 15.

Nehm, R. H. (2018). Evolution. In *Teaching Biology in Schools: Global Research, Issues, and Trends* (pp. 164–177). Routledge.

Nehm, R. H., Beggrow, E. P., Opfer, J. E., & Ha, M. (2012). Reasoning about natural selection: Diagnosing contextual competency using the ACORNS instrument. *The American Biology Teacher*, *74*(2), 92–98.

Nehm, R. H., Poole, T. M., Lyford, M. E., Hoskins, S. G., Carruth, L., Ewers, B. E., & Colberg, P. J. (2009). Does the segregation of evolution in biology textbooks and introductory courses reinforce students' faulty mental models of biology and evolution? *Evolution: Education and Outreach*, *2*(3), 527.

Nehm, R. H., Rector, M. A., & Ha, M. (2010). "Force-talk" in evolutionary explanation: Metaphors and misconceptions. *Evolution: Education and Outreach*, *3*(4), 605.

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*(10), 1131–1160.

Owerkowicz, T., Elsey, R. M., & Hicks, J. W. (2009). Atmospheric oxygen level affects growth trajectory, cardiopulmonary allometry and metabolic rate in the American alligator (*Alligator mississippiensis*). *Journal of Experimental Biology*, *212*(9), 1237–1247.

Sands, D., Parker, M., Hedgeland, H., Jordan, S., & Galloway, R. (2018). Using concept inventories to measure understanding. *Higher Education Pedagogies*, *3*(1), 173–182.

Schank, R., & Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding.* Hillsdale, NJ: Erlbaum.

Scott, E. C. (2010). Dobzhansky was right: Let's tell the students. *Bioessays*, *32*(5), 372–374.

Szabo, M. (1969). The relationship of intellective, personality, and biographical variables to success and its prediction in an independent study science course at the college level. *ERIC*.

Tribune, C. (2017, November). *SAT scores in Illinois: See how your high school compares.* Retrieved from https://www.chicagotribune.com/suburbs/ct-school-report-card-sat-scores-2017-htmlstory.html

UIC. (2020). *UIC Office of Institutional Research.* Retrieved from https://oir.uic.edu/

Vincent, T. L., & Brown, J. S. (2005). *Evolutionary Game Theory, Natural Selection, and Darwinian Dynamics.* Cambridge University Press.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction*, *4*(1), 45–69.

Webb, N. L. (2002). Depth-of-knowledge levels for four content areas. *Language Arts*, *28*.

Ziadie, M., & Andrews, T. (2018). Moving evolution education forward: A systematic analysis of literature to identify gaps in collective knowledge for teaching. *CBE—Life Sciences Education*, *17*(1), ar11.