Transforming a College Biology Course to Engage Students: Exploring Shifts in Evolution Knowledge and Mechanistic Reasoning

Lisa O. Kenyon, Emily M. Walter and William L. Romine

1 Instructional Vignette

1.1 *Environmental Conditions Are Changing and the Birds Are Dying*

Students are about to embark on an online investigation to figure out what is happening to the birds. They clambered onto the only accessible rock to Daphne Major, accompanied by Drs. Peter and Rosemary Grant, 300 miles west of Ecuador. So began their engagement into synthesizing ground finch data from the Galapagos Islands. Students look at large data sets, determine patterns of evidence, and construct explanations about why some finches die and some survived. The students do not always agree. With their partners, students search through the database looking at environmental factors, food availability, predator–prey interactions, and morphometric traits such as weight, wing, beak, and leg length to find evidence that supports their claims about what happened on the Island.

As I walk around the room listening to their discussions, I hear comments such as "I think it has something to do with the rainfall, look [pointing to a graph], the rainfall decreased from 200 cm in wet 1973 to 25 cm in 1977. This was happening the same time the finch population was decreasing, look [pointing to another graph]

L. O. Kenyon (✉) · W. L. Romine
Departments of Biological Sciences and Teacher Education, Wright State University, 3640 Colonel Glenn Highway, Dayton, OH 45435-0001, USA
e-mail: lisa.kenyon@wright.edu

W. L. Romine
e-mail: romine.william@gmail.com

E. M. Walter
Science & Mathematics Education Center, California State University, Fresno 2555 E. San Ramon Ave, Fresno, CA 93740-8034, USA
e-mail: ewalter@csufresno.edu

the population decreased from 60 in wet 1976 to 23 in wet 1977." Another group is arguing that they found evidence that the hawks are eating the finches. "Here, look [pointing at field notes in wet season 1976]. Gf71 was swept up by a hawk, dropped near the waterfront, and devoured."

Students ask me if they have the right answer, checking for my approval. I continually redirect their focus and ask them to think about their evidence; does it support their claim? How could the evidence rebut the claim? Once partners have constructed their explanations, I pair groups to work together converging on one explanation. Students are very chatty; all groups are busy evaluating, modifying, and defending ideas. Students arrive at a variety of explanations, some claim the results are gender-driven, with females out-surviving males, while others are convinced it has to do with beak size and eating the available harder seeds. Students tell me, "Wow, I never really understood natural selection, this makes sense to me know," "We never really learned about evolution in high school," "I never really understood this before because no one explained it this way," and "I totally get this, especially when thinking about how the finch population changed with respect to beak size."

2 Course Overview and Rationale

A central action of many post-secondary pedagogical initiatives is to encourage college and university instructors to adopt approaches based in research on how people learn (AAAS, 2011; Bransford, Brown, & Cocking, 2000). Despite this, efforts to transform the nature of post-secondary instruction have had limited success (e.g., AAAS, 2013), and as many as 70–90% of post-secondary instructors teach exclusively through lecture (Alters, 2005). In these settings, students learn to play the game of memorizing information, but have little or no meaningful learning. These challenges are compounded for topics like evolution, which are difficult to comprehend (e.g., Bishop & Anderson, 1990; Demastes, Good, & Peebles, 1996; Moore et al., 2002) and may be in conflict with students' worldviews.

There are a number of promising strategies for addressing the challenge of effective evolution instruction. The focus of our work in this project was a shift to providing more opportunities for students to build reasoning skills around content knowledge (Berland et al., 2016).

As such, it was our goal to transform an introductory, lecture-based biology course toward a more active learning environment built around science practices. We continued to follow a course content sequence of the typical biology textbook (Reece et al., 2014), but incorporated a different instructional framework based on the Next Generation Science Standards storyline K-12 model (Reiser, 2017). This framework provides a coherent sequence of lessons in which students generate questions by experiencing scientific phenomena. These questions then lead to investigations, situating students in contexts where they figure out problems while engaged in the science practices (NGSS Lead States, 2013).

We foregrounded the importance of evidence as the main objective, thus fostering the use of evidence in figuring out problems through explanatory thinking. We prompted students to answer "how and why" questions through mechanistic responses. Russ, Scherr, Hammer, and Mikeska (2008) define mechanism as a type of causal reasoning addressing how and why individual components of a phenomenon interact with one another over a period of time. More specifically, mechanisms represent non-teleological reasoning (Russ, Coffey, Hammer, & Hutchinson, 2009) and provide the rationale for why a phenomenon occurs. For instance, mechanistic models focus on several particular conditions: target detailed phenomena, identify initial conditions, identify entities, identify the organization of entities, and chain thoughts by working backwards or forwards to explain the situation. Other literature has interpreted mechanisms in similar ways as theoretical accounts that allow for causal explanations and testable predictions about the natural world (Darden & Craver, 2002; Machamer, Darden, & Craver, 2000). Our students collected empirical evidence and used the data to make sense of (a) mechanisms of evolutionary change, (b) body systems, (c) plant biology, and (d) ecology and then connected these phenomena back to the unifying theme of evolution.

3 Literature Review

Like all quality teaching, effective evolution instruction is based at least partly on understanding educational psychology. Much of the theory on how people learn has been translated into various evolution education interventions. At the root of these interventions is the principle that active, evidence-based learning has significant advantages over traditional, lecture-based approaches. In an exhaustive meta-analysis of STEM education research papers, average examination scores improved by about 6% in active learning sections, and students in classes with traditional lecturing were about 1.5 times more likely to fail than were students in classes with active learning (Freeman et al., 2014). A discussion of many of these pedagogical interventions as related to evolution can be found in Andersson and Wallin (2006) and Smith (2010). We summarize some of this literature herein as it fits with our study.

Our intervention model is grounded in two active learning precepts. First, *reasoning and critical thinking are key to building understanding of evolution* (Clough & Wood-Robinson, 1985; Lawson & Weser, 1990; Wandersee, 1985) and acceptance of evolution (Lawson & Worsnop, 1992). Second, *direct experience with phenomena is key to building understanding* of evolution (e.g., Nehm & Reilly, 2007).

3.1 Reasoning and Critical Thinking Is Key to Building Understanding of Evolution

Passmore and Stewart (2002) report on the design and implementation (but not the evaluation) of a nine-week elective high school course on evolution. The goal of this course was to "initiate students into the reasoning patterns of the discipline by engaging them in inquiry contexts that required them to develop, use, and extend Darwin's model of natural selection and to gain some experience with the significance of historical reconstructions" (p. 190). Students examined four real-world data-rich cases using the models of Paley, Lamarck, and Darwin, examining the phenomena to be explained, comparing underlying assumptions/beliefs, and comparing the explanatory power of each. This approach is noted as well grounded and worthy of replication and evaluation (Smith, 2010).

Evaluation of interventions built to improve students' reasoning and critical thinking skills around evolution was a focus of several early evolution education research studies. Much of this work has been done at the primary and secondary levels, but could still apply to a population of college science learners. Lawson and Thompson (1988) argued that formal reasoning skills enable students to modify prior beliefs (e.g., Posner, Strike, Hewson, & Gertzog, 1982), and therefore, the extent to which students hold non-scientific beliefs should be related to this skill. They examined seventh-grade students and found that after receiving instruction on genetics and natural selection, a sample of concrete operational (per Piaget) students held significantly more misconceptions than their formal operational peers. In the case of our study, this implies that if understanding evidence requires formal reasoning skills, it would seem necessary for the students to be formal operational; hence, instruction must be designed to promote its development in concrete operational students. Along these lines, Lawson and Worsnop (1992) noted that skill in reflective reasoning facilitated conceptual knowledge acquisition. They found that grade 10 students who were accomplished reflective (hypothetico-deductive) thinkers exhibited greater conceptual knowledge gains about evolution and natural selection than peers who were less skilled at reasoning.

Lawson and Weser (1990) found that college students who were less skilled at reasoning were more likely to hold non-scientific beliefs and were less likely to change those beliefs during instruction. They also discovered that students who were less skilled at reasoning were also less likely to be strongly committed to scientific beliefs. In other words, students who have poorly developed hypothetico-deductive reasoning skills may hold a correct scientific conception, but may not be strongly committed to that perception. Such students agree with an idea because they have been told that it is correct, rather than arriving at that idea themselves through an internal hypothetico-deductive dialogue around the evidence.

3.2 Direct Experience with Phenomena Is Key to Building Understanding

According to the NGSS (NGSS Lead States, 2013) natural phenomena are "observable events that occur in the universe and that we can use our science knowledge to explain or predict. The goal of building knowledge in science is to develop general ideas, based on evidence, that can explain or predict phenomena." Our intervention model is also supported on the premise that experience with evolutionary phenomena is beneficial to understanding. Scientific ideas are more likely to occur when students can experience phenomena directly (Alters, 2005). Live, eukaryotic organisms with a short generational time are best for observing evolution in action. Experiments using genetically modified foods, *Drosophila* (Coleman & Jensen, 2007; Plunkett & Yampolsky, 2010; Salata, 2002), *E. coli*, or cross-fertilization of plants (Sinatra et al., 2008) provide actual observations of natural, artificial, and/or sexual selection phenomena.

4 Instructional Intervention

As noted in our course design framework (Fig. 1), we integrated multiple activity structures to emphasize empirical reasoning skills and experience with evolutionary phenomena. We first set underlying concepts, anchored by evolution as a unifying concept, and focused on quality rather than quantity of the biological content. We explored questions about the evolutionary origin of animals and plants, their morphology and physiology, and the ecological interactions between organisms and ecosystems they inhabit. Throughout the course, evolution was the unifying theme (e.g., Coker, 2009).

Once we determined weekly topics and associated chapters, we planned for specific activities to teach each day. We selected resources that required students to examine data and use data as evidence to figure out scientific questions and/or make scientific explanations. Based on these criteria, we used Howard Hughes Medical Institute (HHMI) BioInteractive, a free online Web site. HHMI BioInteractive provides a variety of multimedia, apps, videos, interactives, and virtual laboratories that allow students to explore science through a scientific lens. Most media are coupled with student handouts for active learning exercises. These served as formative assessments for the course. We identified four short films, coupled with apps, interactives, and virtual laboratories as contexts that pushed on examining scientific evidence.

In addition to HHMI BioInteractive, we used another computer resource called *Gizmo*. *Gizmos* are online learning simulations that allow students to figure out concepts through making predictions, collecting data, interpreting graphs, and justifying conclusions. We used two *Gizmos* during the class to support student understanding about the digestive and circulatory systems. Other additional computer resources included a Web site called *BGuILE*, The Galapagos Finches, to examine both quan-
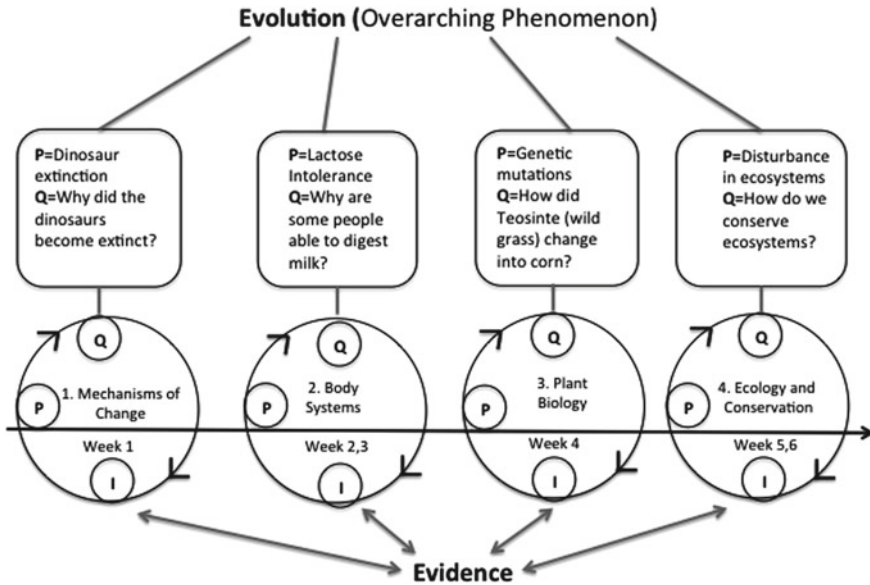
Evolution (Overarching Phenomenon)

Fig. 1 Course design framework. P = phenomenon, Q = questioning, I = investigation. Using HHMI video resources, each unit was anchored with a phenomenon, followed by a question to investigate using evidence. Evolution was the overarching phenomenon that was threaded throughout each unit and provided course coherence. Arguing with evidence was foregrounded as the key practice to figuring out the investigation and used to explain phenomena

titative and qualitative data and explore interrelationships between organisms and environmental influences on a finch population. And lastly, we used a computer simulation program called *NetLogo* to examine changes made to a community of organisms where students were tasked in predicting what happens when we add an unknown invader to this ecological system. In Table 1, we show examples of active learning activities aligned with the lecture material for the course. This table is not a complete list of the activities, but a summary of typical weeks. For each week, we would teach four lectures including 2–3 active learning days. Each activity would include one of the practices (modeling argumentation, or explanation). Students worked in groups during class, making sense of the activity practices or problem-based questions.

To achieve emphasis on empirical reasoning skills and experience with evolutionary phenomena, we developed instructional activities around the NGSS Science and Engineering Practices (SEPs; NGSS Lead States, 2013). While the NGSS was written for a primary and secondary (K-12) audience, the framework could be applied to college classrooms, as there is little reason to believe that high school students learn differently than early college students. Furthermore, both the DBER Report and Vision and Change give scientific practices and content equal importance (AAAS, 2011; NRC, 2012).

Table 1 Examples of lecture material for the course aligned with active learning activities

Lecture material	Active learning activities during course time
History of Life on Earth Mechanisms of Evolution Evolution of Population Origin of Species	Activity 1: HHMI Film *The Day the Mesozoic Died*. Students worked in groups of four to examine the KT boundary trail of *evidence* for why the dinosaurs became extinct. Handouts used in this activity were provided by the HHMI Web site for this particular film Activity 2: BGuILE, Beak of Finch (bird) Investigation. Students worked in pairs to examine a database to construct an explanation for why finches died and why some survived. They used their explanations to argue their claims using *evidence* as they formed consensus explanations with another group pair
Animal Form and Function Animal Nutrition	Activity 3: Gizmo on the Digestive System. Students, working in pairs, use a computer simulation to learn the structure and function of the digestive system. They manipulate the addition and removal of structures to design the best functioning digestive system Activity 4: HHMI film on *Got Lactase? The Co-evolution of Genes and Culture.* Students examine the *evidence* that explains genetic changes associated with the ability to digest lactose in milk, while tracing it to earlier civilizations where human populations started domesticating animals. Handouts used in this activity were provided by the HHMI Web site for this particular film
Circulation Gas Exchange	Activity 5: Students construct scientific models on how and why blood moves through the body. Students share these models in class, participate in a Gizmo to gather *evidence* on circulation, and then revise their models
Evolution of Seed Plants Plant Structure and Growth	Activity 6: HHMI film *Popped Secret: The Mysterious Origin of Corn*. Students, working in pairs, examine the *evidence* of how genetic changes were involved in the transformation of Teosinte (wild grass) into corn. Handouts used in this activity were provided by the HHMI Web site for this particular film

A *scientific practice* represents social and scientific construction, evaluation, and communication of scientific knowledge (Duschl, Schweingruber, & Shouse, 2007). The goal is that students become well-grounded in scientific theory and thus able to form legitimate questions about the natural world around them and then use these practices to discover the answers to their questions. The eight NGSS SEPs include ideas such as *Developing and Using Models*, *Constructing Explanations*, *Planning and Carrying out Investigations*, *Analyzing and Interpreting Data,* and *Engaging in Argument from Evidence*. In the case of our intervention, we engaged students specifically in three NGSS SEPs: (a) developing and using models, (b) constructing explanations, and (c) engaging in an argument from evidence.

In a traditional college course, student engagement with SEPs likely would be relegated to the laboratory. This is a missed opportunity from our perspective. Therefore, our ancillary goal was to transform a lecture environment using technological tools for social sense making around the SEPs. We introduced "*untethering*," a process that begins with mobile device mirroring with a tablet (in this case, an iPad). The

instructor uses the iPad as a tool to untether from the podium and walk around the room to engage students in the material and discussions. The use of the iPad device provided opportunities for leveraging inquiry-based apps that allowed the professor and students to share student work and ideas for whole class discussions. More students participated in peer-collaborative learning through the technology with this pedagogical initiative (Thinley, Reye, & Geva, 2014).

5   Research Methods

5.1   Paradigm

We assume a post-positivist paradigm for this research study, reflecting a single, objective reality that is measurable by survey data. We therefore chose to ask research questions that fit a quantitative approach to approximate a single reality, i.e., "What do the students know?" Post-positivism as a paradigm challenges the traditional, positivist idea of an absolute truth (Phillips & Burbules, 2000) and recognizes that we cannot always know reality when studying behavior and actions of human subjects. Reality from a post-positivist view is based on cautious observation and measurement of the objective reality that exists "out there" in the world. In our case, exploring the thinking of individuals through survey data reflects our post-positivist paradigm (Creswell, 2009).

5.2   Context

We examined 70 science majors in an introductory biology class at a research-intensive, open-enrollment university in the Midwest USA. The total minority student enrollment was 19.7% (10.4% African-American, 3.3% two or more races, 2.8% Asian American, 2.9% Hispanic American, 0.2% American Indian or Alaskan Native, and 0.1% Native Hawaiian or Pacific Islander), and the total international student enrollment was 11% (65 countries). As this class was an introductory course, its demographic distribution was consistent with the university as a whole. The course took place during a summer term when students met for six weeks, four days a week, for 1.5 h a session. Most students were science majors (many pre-medical profession), engineering, and computer science majors and had taken the prerequisite course on cells and genetic biology. Few, if any, students experienced active learning in their prior courses since attending the university. Typically, this summer session class was lecture only. These longer-time sessions allowed for this redesign opportunity to engage students in practices that support both cognitive and social approaches to the scientific discipline. The course professor had a graduate degree in biological sciences and, most influentially, had a doctorate in science education with

an emphasis on curriculum design and instruction. The professor used her research and teaching philosophy to guide the design modifications in this course.

5.3 Instrumentation

5.3.1 Knowledge of Natural Selection

We used the Conceptual Inventory of Natural Selection (CINS; Anderson, Fisher, & Norman, 2002) as a measure for knowledge of natural selection before and after instruction. The CINS was developed in response to previous instruments (Bishop & Anderson, 1990; Settlage & Odom, 1995) because the authors found the old instruments to be overly simplistic and abstract. Their solution to this was to develop an instrument that used actual evolutionary examples (e.g., Galapagos finches, Venezuelan guppies *Poecilia reticulata*, and Canary Island lizards). The 20-item CINS was therefore developed to measure non-science majors' understanding of natural selection. It was designed for each item to have one correct answer and three distracter answers based on common alternative conceptions about natural selection. The questions on the CINS target seven key concepts of natural selection (Mayr, 1982) and two additional key concepts (origin of variation and origin of species). Two questions target each key concept to enhance reliability. In the context of this study, selection refers to: causes of phenotypic variation (e.g., mutation, recombination, sexual reproduction); heritability of phenotypic variation; the over-reproductive capacity of individuals; limited environmental resources or carrying capacity; competition or limited survival potential; selective survival based on heritable traits; and changes in the frequency of individuals with certain heritable traits (Mayr, 1982, pp. 479–80).

Prior research with the CINS demonstrates validity and reliability sufficient for group or temporal comparisons (reliability > 0.7). In this study, we used the CINS as a measure of a single latent variable (knowledge of natural selection) in line with the validity analysis of Anderson et al. (2002). When used in this way, the CINS demonstrated satisfactory reliability (Rasch reliability = 0.75) in our sample of college students and all items demonstrated satisfactory weighted mean squares fit with the Rasch validity model (Wright & Stone, 1979).

5.3.2 Mechanistic Reasoning

We used a single constructed response assessment item to solicit mechanistic reasoning (Krist, Schwarz, & Reiser, 2018). In response to the prompt, students were asked to specify the factors they believed contributed to increased percentage of elephants without tusks and were subsequently asked to *explain their reasoning* behind their response. Students responded to the following prompt:

- African elephants are known for their large tusks, which the animals use for digging and defense. These tusks are valuable to people because of their ivory, which can be used in jewelry and decorations. Poachers hunt and kill elephants for their tusks, often before elephants are able to reproduce. Some elephants never grow tusks. In 1930, 1% of adult elephants didn't have tusks. In some areas today, up to 38% of adult elephants don't have tusks.
- How and why is the percentage of elephants without tusks higher today than it was in 1930?

5.3.3   Student Assessment of Learning Gains

At the end of the semester, we administered an online survey called "Student Assessment of their Learning Gains" (SALG) to measure students' self-reported learning gains and other progress toward course learning outcomes. The survey consisted of a variety of constructs including content understanding, skills, attitudes, class activities, class resources, and student support. Student responses were reported to the instructor after the course was completed. Aligning with our research questions, we report student survey data regarding content understanding and increase in skills.

6   Research Questions

6.1   How Did Students' Understanding of Natural Selection Change During the Course?

We sought interpretations of how concepts about natural selection changed, which involved: (1) statistical significance of gains; (2) changes in conceptions implied by the gains; and (3) students' assessment of their learning gains. First, we were interested in whether knowledge of natural selection improved. To aid interpretation, Rasch logit measures on the CINS were first rescaled onto a range of 0–20 (the range for the original CINS scale). Change in the mean measure before and after the class was evaluated at the 0.05 alpha level using a paired t test. Since the distributions were not normal, standard errors, confidence intervals, and *p* values were derived from a bootstrap distribution based on 10,000 simple random draws with replacement from the data. We used the percentile method to generate a 95% confidence interval for gains (0.025 and 0.975 quantiles of the bootstrap distribution) (Banjanovic & Osborne, 2016). The standardized mean gain (Cohen's D) was used as a measure of practical significance. Cohen's (1988) guidelines were used to qualify the size of the effect from the standardized mean difference.

Second, we constructed a Wright map of student and item Rasch measures along the common CINS scale (Fig. 2). The Wright map is a plot of student and item measures along a common scale and allows one to predict concepts that individual students have mastered based on their relative location along the scale (Boone, 2016).
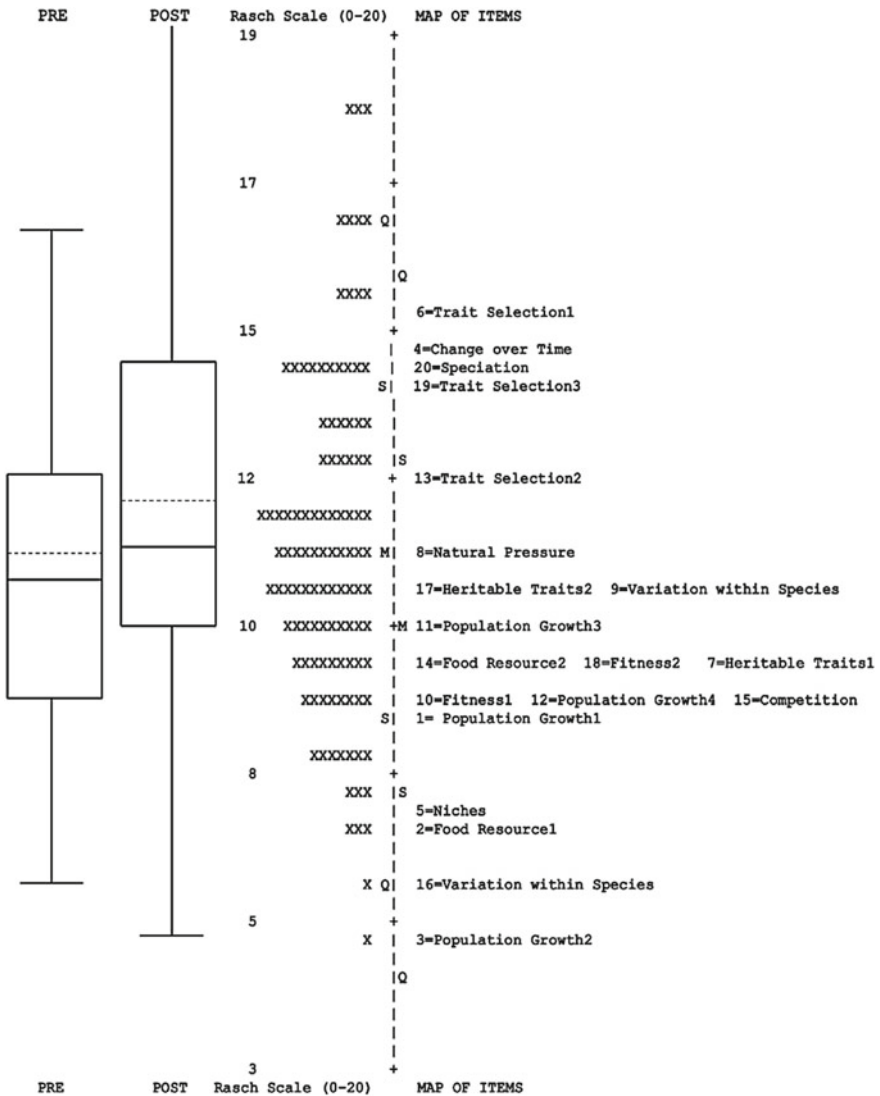

Fig. 2 Wright map of student and item measures along the CINS scale (0–20). Students who are positioned below the location of the item are predicted to get that item incorrect, indicating non-mastery of that concept. The x's show the total distribution of student measures (pre and post together). The box plots indicate the distribution of students' measures before and after instruction

Specifically, if a student's ability location sits above the item's difficulty location on the scale, then that student is predicted to have mastered the concept associated with that item. From the Wright map, we were able to deduce concepts that were comparatively easy or difficult for students, and how mastery of particular concepts changed between the beginning and end of instruction. All analyses were carried out under the assumption that interpretation of measures on the CINS did not change between the pre- and posttest.

6.2 How Did Students' Mechanistic Reasoning Around Natural Selection Change?

We scored the mechanistic reasoning prompt on two levels. For level 1, we coded students' responses based on factors they believed caused more elephants without tusks, "within variation in a trait or genes exists within a population of organisms" (1a in Table 2) and "humans caused a change in the environment which selected for elephants without tusks" (1b in Table 2). We also were interested in whether students provided inaccurate alternative explanations (1c in Table 2). For level 2, after students described the factors, we asked them to explain their reasoning. If students reasoned that variation in traits or genes (1a) happens because organisms reproduce and pass on genes or traits to offspring, then they were deemed to show appropriate reasoning (2a in Table 2). Similarly, appropriate reasoning around natural selection (1b) involved explanation that human hunting affected the elephant population over time (2b in Table 2).

We also documented responses containing reasoning behind misconceptions about how other processes may have caused the elephant population to change (2c in Table 2). We hypothesized that effective instruction would increase the proportion of students who specified the correct factors causing the change in the elephant population (increase in 1a and 2a), and correct reasoning around how these factors caused the change (increase in 2a and 2b). Since all six of these tasks were scored dichotomously (and hence the distributions were not normal), standard errors, confidence intervals, and p values were again derived from a bootstrap distribution based on 10,000 simple random draws with replacement from the data. We used the percentile method to generate 95% confidence intervals for gains (0.025 and 0.975 quantiles of the bootstrap distribution) (Banjanovic & Osborne, 2016). The standardized mean gain (Cohen's D) was used as a measure of practical significance of student gains or losses before and after the unit.

Table 2 Changes in students' concept understanding and mechanistic reasoning around natural selection through the course

Construct	Pre	Post	P value	95% CI_{Gain}	D_{gain}
Natural selection[a]	10.83 (2.49)	11.83 (3.06)	0.011	0.29 to 1.74	0.39*
Variation 1a[b]	0.22 (0.42)	0.32 (0.47)	0.31	−0.08 to 0.28	0.15
Selection 1b[b]	0.80 (0.40)	0.78 (0.42)	0.79	−0.16 to 0.12	−0.04
Inaccurate 1c[b]	0.58 (0.50)	0.38 (0.49)	0.026	−0.36 to −0.04	−0.33*
Correct reasoning 2a[b]	0.06 (0.24)	0.24 (0.43)	0.017	0.04 to 0.32	0.34*
Correct explanation 2b[b]	0.54 (0.50)	0.44 (0.50)	0.214	−0.26 to 0.04	−0.18
Incorrect reasoning 2c[b]	0.44 (0.50)	0.26 (0.44)	0.049	−0.34 to 0.00	−0.29*

*Effect size is significant at 0.05 alpha level
[a]Paired t test for difference in means ($N = 46$) reported as Mean (SD). Cohen's D (D_{gain}) used as an effect size measure. Cohen's D (D_{gain}) was used as an effect size measure. Since this distribution is not normal, 95% confidence intervals and 2-tailed p values were derived using the percentile method (0.025 and 0.975 quantiles of the bootstrap distribution) using 10,000 simple random draws with replacement from the data
[b]Paired t test for difference in means ($N = 50$) reported as Mean (SD). Cohen's D (D_{gain}) was used as an effect size measure. Since these distributions are not normal, 95% confidence intervals and 2-tailed p values were derived using the percentile method (0.025 and 0.975 quantiles of the bootstrap distribution) using 10,000 simple random draws with replacement from the data

7 Findings and Discussion

In this section, we present our findings and relate them to the literature. We focus on data from the CINS and our mechanistic reasoning prompt. We conclude with pedagogical implications and a reflection on the course (including SALG comments from students), as well as implications for future research and faculty development.

7.1 Overall Shifts in Natural Selection Knowledge

Students' conceptions about evolution were more sophisticated by the end of the course (bootstrap 95% CI = 0.29–1.74, $p = 0.011$, $D_{gain} = 0.39$). A standardized mean difference of 0.39 is equivalent to moving from the 50th percentile to the 65th percentile in average performance on the CINS. We found that this gain was accompanied by a proportional decrease (bootstrap 95% CI = −0.36 to −0.04, $p = 0.026$, $D_{gain} = 0.33$) in inaccurate alternative explanations (Inaccurate 1c in Table 2). Both these findings concord with Nehm and Reilly (2007), who also investigated science majors' natural selection knowledge and alternative conceptions in an active

learning setting. Nehm and Reilly quantified individual use of key concepts versus alternative conceptions into a single composite measure called the natural selection performance quotient (NSPQ). A passing NSPQ was 65 (out of 100), a score calibrated to require employment of at least four of Mayr's seven key concepts (1982). As in our study, knowledge of natural selection was low prior to instruction (62; failing). Post-course, Nehm and Reilly documented a significant knowledge increase in their active learning group (from 62 to 79).

7.2   Item-Level Shifts in Natural Selection Knowledge

In Fig. 2, we document differences in mastery of concepts before and after instruction based on our CINS data. As it is common for students to show extremes on one side or another, we were most interested in the middle of the distribution; namely how the first, second (median), and third quartiles of the distributions changed, and the subsequent inferences we can draw with respect to concept mastery.

The median (50th percentile) of the student measure distribution (solid line in the middle of each box) shifted from 10.5 to 11.1. Item 8 sits between these two measures, indicating that a median student did not understand the role of the environment in selecting for certain beak types in Darwin's finches before the class, but that they gained this understanding by the end of the class. The first quartile (25th percentile) of the distribution shifted from 8.6 to 9.9 between the pre- and posttests. Items 7, 14, and 18 sit between these levels, indicating that students at the second quartile obtained mastery of the concepts of (a) heritability, (b) competition for resources, and (c) fitness. Furthermore, mastery of item 7 indicates that the course helped these students abandon the Lamarckian misconception that change occurs due to a need or desire. This was replaced by the understanding that genes are a driver of evolution. The students also understood the biological definition of "fitness" by the end of the class (item 18) and expressed understanding that resource limitations exist (item 14).

The third quartile (75th percentile) of the distribution shifted from 12.4 to 14.1, indicating mastery of items 19 and 20. These items relate to selection of traits and speciation. Item 19 indicates that the unit may have helped students replace Lamarckian misconceptions of within-species phenotypic variation with the understanding that random genetic mutations are the initial driver of variation, and item 20 indicates that students were then able to apply this idea toward scientifically accepted understanding of how speciation occurs.

7.3   Changes in Students' Mechanistic Reasoning

We originally hypothesized that effective instruction would increase the proportion of students with correct reasoning around how natural selection caused changes (items 2a and 2b). Students' ability to qualify the variability in genes and traits in a

population did not change through the course (Item 2b), but their ability to explain the reasoning behind this—that organisms pass genes and traits to offspring (item 2a)—did increase significantly (bootstrap 95% CI = 0.04 to 0.32, p = 0.017, D_{gain} = 0.34). Only 6% of students could express this reasoning clearly at the beginning of the course; this increased to 24% of the students by the end of the course. While this is not the level of reasoning mastery we would like to see in our science major students, we see this as a step in the right direction.

7.4 Pedagogical Implications and Reflections on the Course

Our goals were to present students with phenomena and engage them in using evidence to explain and reason through the phenomena. Our data indicate that the course was effective in transforming students' conceptions about evolution. In particular, we found that the introductory investigation of how and why the dinosaurs died was pivotal in changing the climate and mindset of the course. We showed this HHMI video on the first day and students worked in small groups figuring out the evidence of what happened to the dinosaurs. It was clear that the students were excited and felt that this course was going to be "different" than other courses. For example, on the SALG survey student comments included:

- "The class activities were awesome! They helped me so much to understand the material—I would even go home and talk about what I learned to others because I found it very interesting and exciting!"
- "I liked looking back and applying what we learned at the start of the course and building up and growing upon the idea of evolution and branching from there."

When students learn about evolution, it is often through direct instruction, personal experiences, and or in bits and pieces, which may lead to misconceptions and incomplete understandings (Coker, 2009; Gil-Perez & Carrascosa, 1990; Sinclair, Pendarvis, & Baldwin, 1997). Decontextualized experiences may explain in part why students experience difficulty when learning evolution. We attribute our relative success to the framing of the course and the relevancy of using the situated context for learning about the overarching phenomenon of evolution. We chose the history of life on Earth as an entry point to thinking about evolution, particularly understanding the adaptive radiation of mammals. The film about dinosaur extinction allowed for this content to become more interesting. Most students have heard about the dinosaur extinction, dating back to their preschool and elementary years, but this topic rarely appears in their upper science classes. This context opened up the space for bringing in different student ideas of what happened, which led to a variety of questions to investigate. Thus, students had a personal interest in the topic.

Our focus on foregrounding evidence was key for facilitating student buy-in and understanding and explaining evolution. The course message was not about finding the right answer, but rather examining the evidence to figure out the most convincing claim. Changing the language of the classroom environment to include attention to

audience and persuasion with evidence contributed to a positive change in student thinking and learning. When students were asked what skills they learned from participation in this course, they commented, "I learned how to learn," "Being able to analyze different pieces of evidence and putting that together, like what explained the K-T extinction" and "One of the main skills that I have gained as a result of this class is looking at both sides of an argument. I feel like I tend to always pick one side but never really look on the other side of the argument. This class really challenged me to analyze both sides of the argument and actually find evidence to 'support the claim'."

The film also provided a look into the personal side of science, where scientists agreed or disagreed with one another. Finally, we used technology as a tool throughout the course to provide opportunities for student learning. Student comments on the SALG included, "The technology was extremely helpful, and how she tied in many different forms of learning. I do not necessarily learn well from just being lectured at, but rather we watched videos and did interactive gizmos, which extremely helped." and "I really appreciated the great efforts in use of technology in the class, it was great having such a forward thinking professor."

7.5 Research Implications

We encourage other researchers to explore mechanistic reasoning behind evolutionary concepts. We document gains in reasoning (Table 1) and strong indications that the science practice approach positively influenced student understandings of natural selection. In particular, we are interested in exploring mechanistic reasoning more comprehensively—and exploring not only reasoning behind not natural selection mechanisms, but also those related to speciation. This approach addresses the call for students to have a complete scientific understanding of evolution. They should learn examples of natural selection and speciation on both microevolutionary and macroevolutionary scales (Catley, 2006). Evolution across long timescales may be particularly important as knowledge of macroevolution has been reported to be significantly correlated with acceptance of evolution for both biology (Nadelson & Southerland, 2010) and non-science majors (Romine, Walter, Bosse, & Todd, 2016; Walter, 2013; Walter, Halverson, & Boyce, 2013).

7.6 Implications for Faculty Development

Instructors often create lessons, select readings, and design assessments in the same way they always have (Wilson, 2010), calling on their experiences as learners to inform how they teach (Tobin, Tippins, & Gallard, 1994). In this way, instructors can perpetuate ineffective and antiquated lecture norms as they operate under the belief that teaching occurs by transmitting knowledge (DeHaan, 2005).

We deviate from lecture-only approaches in our intervention, as we are using phenomena to help students engage in authentic science practices. Unlike other studies that incorporate principles of inquiry in this manner (e.g., Demastes, Settlage, & Good, 1995; Robbins & Roy, 2007), our intervention occurs in a large enrollment lecture hall, not a laboratory classroom. Since the SEPs model how scientists understand and practice in their own work, implementation of a practice-based teaching approach may provide an easier pedagogical transition for faculty new to active learning strategies. For example, we postulate that a faculty member unsure on how to implement an approach like "problem-based learning" may feel more comfortable with guiding students to "build an argument from evidence." In this way, our study could be used as a bridge between these two worlds: how instructors teach and how students learn.

References

Abraham, J. K., Meir, E., Perry, J., Herron, J. C., Maruca, S., & Stal, D. (2009). Addressing undergraduate student misconceptions about natural selection with an interactive simulated laboratory. *Evolution: Education and Outreach*, *2,* 393–404.

Alters, B. T. (2005). *Teaching biological evolution in higher education: Methodological, religious, and nonreligious issues*. Sudbury, MA: Jones and Bartlett Publishers.

American Association for the Advancement of Science [AAAS]. (2011). *Vision and change in undergraduate biology education: A call to action.* Washington, DC: Author.

American Association for the Advancement of Science [AAAS]. (2013). *Measuring STEM teaching practices: A report from a national meeting on the measurement of undergraduate Science, Technology, Engineering, and Mathematics (STEM) teaching*. Washington, DC: Author.

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.

Andersson, B., & Wallin, A. (2006). On developing content-oriented theories taking biological evolution as an example. *International Journal of Science Education, 28,* 673–695.

Banjanovic, E. S., & Osborne, J. W. (2016). Confidence intervals for effect sizes: Applying bootstrap resampling. *Practical Assessment, Research & Evaluation, 21*(5), 1–20.

Berland, L. K., Schwartz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching, 53,* 1082–1112.

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.

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE-Life Sciences Education*, *15*(4), rm4.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn*. Washington, DC: National Academies Press.

Campbell, N., & Reece, J. (2012). *Biology* (10th ed.). San Francisco, CA: Pearson Benjamin Cummings.

Catley, K. M. (2006). Darwin's missing link—A novel paradigm for evolution education. *Science Education, 90,* 767–783.

Clough, E., & Wood-Robinson, C. (1985). How secondary students interpret instances of biological adaptation. *Journal of Biological Education, 19*(2), 125–130.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Coker, J. S. (2009). Reinventing life: Introductory biology for a rapidly evolving world. *The American Biology Teacher, 71,* 281–284.

Coleman, S. W., & Jensen, M. S. (2007). Male mating success: Preference or prowess? Investigating sexual selection in the laboratory using *Drosophila melanogaster*. *The American Biology Teacher, 69,* 351–358.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Los Angeles, CA: SAGE.

Darden, L., & Craver, C. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 33*(1), 1–28.

DeHaan, R. (2005). The impending revolution in undergraduate science education. *Journal of Science Education and Technology, 14,* 253–269.

Demastes, S. S., Good, R. G., & Peebles, P. (1996). Students' conceptual ecologies and the process of conceptual change in evolution. *Science Education, 79,* 637–666.

Demastes, S. S., Settlage, J., & Good, R. (1995). Students' conceptions of natural selection and its role in evolution: Cases of replication and comparison. *Journal of Research in Science Teaching, 32,* 535–550.

Detterman, D. K., & Sternberg, R. J. (1993). *Transfer on trial: Intelligence, cognition, and instruction*. Westport, CT: Ablex Publishing.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.

Evans, J. H., & Evans, M. S. (2008). Religion and science: Beyond the epistemological conflict narrative. *Annual Review of Sociology, 34,* 87–105.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., et al. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111,* 8410–8415.

Gil-Perez, D., & Carrascosa, J. (1990). What to do about science "misconceptions". *Science Education, 74,* 531–540.

Krist, C., Schwarz, C. V., & Reiser, B. J. (2018). Identifying essential epistemic heuristics for guiding mechanistic reasoning in science learning. *Journal of the Learning Sciences*. https://doi.org/10.1080/10508406.2018.1510404.

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.

Lawson, A. E., & Weser, J. (1990). The rejection of nonscientific beliefs about life: Effects of instruction and reasoning skills. *Journal of Research in Science Teaching, 27,* 589–606.

Lawson, A. E., & Worsnop, W. A. (1992). Learning about evolution and rejecting a belief in special creation: Effects of reflective reasoning skill, prior knowledge, prior belief and religious commitment. *Journal of Research in Science Teaching, 29*(2), 143–166.

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex Publishing.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science,* 1–25.

Mayr, E. (1982). *The growth of biological thought*. Cambridge, MA: Harvard University Press.

McKeachie, W. J., Lin, Y. G., & Strayer, J. (2002). Creationist vs. evolutionary beliefs: Effects on learning biology. *The American Biology Teacher*, *64*(3), 189–192.

Moore, R., Mitchell, G., Bally, R., Inglis, M., Day, J., & Jacobs, D. (2002). Undergraduates' understanding of evolution: Ascriptions of agency as a problem for student learning. *Journal of Biological Education, 36,* 65–71.

Nadelson, L. S., & Southerland, S. A. (2010). Examining the interaction of acceptance and understanding: How does the relationship change with a focus on macroevolution. *Evolution: Education and Outreach, 4,* 82–88.

National Research Council [NRC]. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies 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. (2007). Does increasing biology teacher knowledge of evolution and the nature of science lead to greater preference for the teaching of evolution in schools? *Journal of Science Teacher Education, 18,* 699–723.

Nelson, C. E. (2008). Teaching evolution (and all of biology) more effectively: Strategies for engagement, critical reasoning, and confronting misconceptions. *Integrative and Comparative Biology, 48,* 213–225.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: The National Academies Press.

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

Phillips, D. C., & Burbules, N. C. (2000). *Post-positivism and educational research*. Lanham, MD: Rowman & Littlefield.

Plunkett, A. D., & Yampolsky, L. Y. (2010). When a fly has to fly to reproduce: Selection against conditional recessive alleles in *Drosophila*. *The American Biology Teacher, 72*(1), 12–15.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education*, *66,* 211–227.

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., Jackson, R., et al. (2014). *Campbell biology* (10th ed.). Boston, MA: Pearson.

Reiser, B. J. (2017). *Developing coherent storylines to support three-dimensional science learning*. Presented at Rhode Island Science Teachers Association (RISTA) Conference.

Robbins, J. R., & Roy, P. (2007). The natural selection: Identifying & correcting non-science student preconceptions through an inquiry-based, critical approach to evolution. *The American Biology Teacher, 69,* 460–466.

Romine, W. L., Walter, E. M., Bosse, E., & Todd, A. (2016). Understanding patterns of evolution acceptance—A new implementation of the Measure of Acceptance of the Theory of Evolution (MATE) with midwestern college students. *Journal of Research in Science Teaching, 54*(5), 642–671.

Russ, R. S., Coffey, J. E., Hammer, D., & Hutchison, P. (2009). Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education, 93*(5), 875–891.

Russ, R. S., Scherr, R. E., Hammer, D., & Mikeska, J. (2008). Recognizing mechanistic reasoning in student scientific inquiry: A framework for discourse analysis developed from philosophy of science. *Science Education, 92,* 499–525.

Salata, M. (2002). Evolution lab with *Drosophila*. *Bioscene, 28*(2), 3–6.

Settlage, J., & Odom, A. L. (1995, April). *Natural selection conceptions assessment: Development of the two-tier test "Understanding Biological Change"*. Paper presentation at the National Association of Research in Science Teaching Annual Meeting, San Francisco, CA.

Sinatra, G. M., Brem, S. K., & Evans, E. M. (2008). Changing minds? Implications of conceptual change for teaching and learning about biological evolution. *Evolution: Education and Outreach*, *1,* 189–195.

Sinclair, A., Pendarvis, M. P., & Baldwin, B. (1997). The relationship between college zoology students' beliefs about evolutionary theory and religion. *Journal of Research and Development in Education, 30,* 118–125.

Smith, M. U. (2010). Current status of research in teaching and learning evolution: II. Pedagogical issues. *Science & Education, 19,* 539–571.

Thinley, P., Geva, S., & Reye, J. (2014). Tablets (iPad) for M-learning in the context of social constructivism to institute an effective learning environment. *International Journal of Interactive Mobile Technologies, 8*(1), 16–20.

Tobin, K. G., Tippins, D. J., & Gallard, A. J. (1994). Research on instructional strategies for teaching science. In D. L. Gabel (Ed.), *Handbook of research on science teaching and learning* (pp. 45–93). New York, NY: Macmillan.

Walter, E. M. (2013). *The influence of pedagogical content knowledge (PCK) for teaching macroevolution on student outcomes in a general education biology course* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses (#3576041).

Walter, E. M., Halverson, K. L., & Boyce, C.-J. (2013). Investigating the relationship between college students' acceptance of evolution and tree thinking understanding. *Evolution: Education and Outreach, 6,* 26.

Wandersee, J. (1985). Are there too many terms to learn in biology? *The American Biology Teacher, 47,* 346–347.

Wilson, R. (2010, September). Why teaching is not priority no. 1. *Chronicle of Higher Education, 57,* A1–A8.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: University of Chicago Social Research.

Dr. Lisa Kenyon is Associate Professor in the Department of Biological Sciences in the College of Science and Mathematics and the Department of Teacher Education in the College of Education and Human Services at Wright State University. Her research focuses on engaging K-12 students and teachers in scientific practices such as explanation, argumentation, and scientific modeling—specifically, examining how students use their epistemologies of science to support these practices. Other research interests include curriculum design, project-based inquiry, and teacher professional development.

Emily M. Walter is Assistant Professor of Biology at California State University—Fresno. At Fresno State, she is Director of the Science and Mathematics Education Center, a group of faculty and students that collaborate across STEM departments and with the School of Education to support STEM education initiatives and research at all grade levels. Her areas of expertise include evolution education and exploring the cognitive, social, and organizational factors that influence faculty adoption and sustained use of evidence-based instructional practices.

William Romine is Associate Professor in Wright State University's Biological Sciences department, with joint appointments in Teacher Education and Computer Science. He directs the *Data Science for Education* laboratory at Wright State where he and his students focus on educational measurement using traditional tests and surveys, social media, and sensors. He and his students use these measures to construct dynamic models for both learning and affective change related to science.