Developing & Testing Curricula for Teaching Evolutionary Concepts at the Elementary School Level

GREGORY F. GRETHER

Abstract

Evolution by natural selection is key to understanding life and of considerable practical importance in public health, medicine, biotechnology, and agriculture. The Next Generation Science Standards (NGSS) include natural selection among several evolutionary concepts that all third-graders should know. This article explores a novel approach to developing and testing curricula for teaching natural selection and related concepts to children. College students developed lesson plans with specific evolutionary learning objectives based on the NGSS and taught them at elementary schools. Learning was assessed with a pre/post-test design, and a subset of students was retested after two years. After just two hours of instruction and active-learning activities, students of all three grade levels tested (grades 3–5) demonstrated substantial improvement in their understanding of evolutionary concepts. Students who were retested in grade 5 scored higher than fifth-graders who had not participated previously. The most challenging concepts for all grade levels were common ancestry and natural selection, but fifth-graders showed more improvement than third- and fourth-graders. If this finding is substantiated by further research, an adjustment to the NGSS schedule might be warranted. Spacing evolutionary biology concepts out might be a better strategy than concentrating them all in grade 3.

Key Words: *adaptation; common ancestry; evolution; evolutionary time; fossils; natural selection; Next Generation Science Standards; NRC framework; science education; vestigial traits.*

○ Introduction

Evolution is among the greatest of human discoveries and key to understanding everything biological (Dobzhansky, 1973). An understanding of natural selection, in particular, is valuable in numerous industries and sectors of the economy, including public health, medicine, biotechnology, resource management, and agriculture, and yet a large proportion of adults do not understand how it works (Gregory, 2009). People who understand natural selection generally regard it as simple, intuitive, and inevitable, but there is ample evidence that it is actually a difficult concept to grasp and

that misconceptions developed in childhood can be difficult to correct later (reviewed in Gregory, 2009; Prinou et al., 2011; Emmons et al., 2018; Lucci & Cooper, 2019). Children naturally grapple with biological questions and deserve to be introduced to evolutionary concepts at a young age, before unscientific ideas become too deeply engrained (Nadelson et al., 2009; Emmons et al., 2018). Providing children with concrete evidence for evolution, such as fossils and vestigial traits, can help counteract cognitive and cultural biases against evolutionary thinking (Evans, 2000; Hermann, 2011). Children of elementary school age are also quite capable of understanding the building blocks of natural selection – within-species variation, mutation, heritability, differential survival, and reproduction – as well as the cornerstone concept of adaptation (e.g., Nadelson et al., 2009; Campos & Sá-Pinto, 2013; Emmons & Kelemen, 2015).

In 2012, the National Research Council (NRC) published an influential report outlining a new framework for K–12 science education. One of the guiding principles of the NRC framework is that evolution and natural selection are "key to understanding both the unity and the diversity of life on Earth" (National Research Council, 2012). The *Next Generation Science Standards* (NGSS) were largely based on the NRC framework and have thus far been adopted by 20 U.S. states (NGSS Lead States, 2013; National Science Teachers Association, 2019). The NGSS are framed in terms of what all students of a given grade level should know about and be able to do to demonstrate their knowledge. According to the NGSS, by grade 3 (i.e., eight to nine years of age), children should know about trait variation and inheritance, fossils and extinct organisms, common ancestry, biological diversity, natural selection, and adaptation (California Department of Education, 2019).

In California, the NGSS were approved for implementation in 2017, but meeting these new science standards is a major challenge. Most elementary schools lack science teachers, and many teachers lack the time or knowledge to design lessons based on the new standards (Dorph et al., 2007; Watanabe, 2011). While the previous elementary school science standards included several of the building-block concepts, they did not include natural selection itself (California Department of Education, 2004). Ensuring that accurate, effective, and easy-to-implement curricula for teaching

evolutionary concepts to children are readily available to elementary school teachers is crucial for the success of the NGSS (Krajcik et al., 2014; Anderson et al., 2018; Lucci & Cooper, 2019).

Which evolutionary concepts do children struggle with the most? What are the most effective ways to teach evolutionary concepts at the elementary school level? How much time, in a regular classroom setting, needs to be devoted to these concepts for students to grasp and retain them? Is grade 3 optimal for introducing common ancestry and natural selection? Or would it be better to introduce these topics later in elementary school, as originally proposed by the NRC (National Research Council, 2012)? With the goal of helping to answer these questions, here I explore a new approach to developing and testing curricula for teaching evolutionary concepts at the elementary school level. The overall design of the study was for college students to develop lesson plans with specific evolutionary learning objectives based on the Disciplinary Core Idea (DCI) dimension of the NGSS and teach them at local elementary schools. Short-term learning was assessed by administering quizzes before and after the lessons. Long-term learning was assessed by retesting a subset of the students two years later and comparing their quiz scores to those of students at the same grade level who had not participated previously. Examples of effective lesson plans are provided with a companion article in this issue of *ABT* (Grether et al., 2021).

○ Methods

Participants

A written application was used to identify University of California, Los Angeles (UCLA), undergraduates with good qualifications and motivations for participating in the study; 148 students applied and 43 were selected to participate. The selected students had taken courses in evolutionary biology, were highly motivated to obtain teaching experience, and were able to outline suitable topics for teaching evolutionary concepts to children.

Two public elementary schools in Los Angeles County participated in both years of the study (2016, 2018). At "school A," all classes in grades 3, 4, and 5 participated (11 classes, 234 students). Most fifth-graders who participated in 2018 (n = 48) had also participated when they were in grade 3 (n = 37). At "school B," grades 3 and 4 were combined and all classes at that level participated in the study (10 classes, 233 students). At school A, third-graders ranged in age from eight to 10 years (mean ± SD = 8.65 ± 0.48; n = 98), fourth-graders ranged in age from nine to 11 years (9.69 ± 0.52; n = 74), and fifth-graders ranged in age from 10 to 12 years (10.67 ± 0.50; n = 73). At school B, students ranged in age from eight to 10 years (9.03 ± 0.72; n = 219).

Development of Lesson Plans

Prior to developing lesson plans, the undergraduates read and discussed articles on evolutionary concepts and misconceptions (e.g., Baum et al., 2005; Nadelson et al., 2009; Grether, 2010a, b; Prinou et al., 2011; Campos & Sá-Pinto, 2013; Padian, 2013; Young et al., 2013; Mervis, 2015) and the effectiveness of different teaching methods (e.g., Kirschner et al., 2006; Hmelo-Silver et al., 2007; Clark et al., 2012; Rosenshine, 2012), and studied relevant sections of the NRC framework (National Research Council, 2012) and the NGSS. They were also provided with links to websites about teaching evolutionary concepts and articles on the local Pleistocene fauna (to improve their understanding of the fossils in the teaching collection; e.g., Carbone et al., 2009; Binder & Van Valkenburgh, 2010; Ripple & Van Valkenburgh, 2010). They worked in small groups (two or three students) to develop lesson plans based on six specific

Table 1. Evolutionary biology learning objectives that served as the target for lesson plans in this study.

1.	**Fossils**	Fossils are organisms that lived long ago; they can show us how organisms have evolved over time and how groups of modern organisms are related to each other; fossils can also tell us about past environments; fossils show us that life has been evolving on Earth for at least 3,500,000,000 (3.5 billion) years.
2.	**Vestigial traits**	Vestigial (useless, leftover) traits provide clear evidence of common ancestry and descent with modification (i.e., evolution).
3.	**Common ancestry**	All organisms are related to each other because they evolved from a common ancestor; evolution is a branching process (tree) not a progression (ladder); all organisms alive today are equally highly evolved.
4.	**Heritability**	Most traits of organisms are variable, and some of the variation is heritable (genetic) and can be passed from one generation to the next.
5.	**Natural selection**	Evolution by natural selection happens because variation in heritable traits affects survival and reproduction; organisms are adapted to their natural environment because of natural selection in the past.
6.	**Evolutionary time**	Evolution in nature is very slow or at least seems slow to us. It takes many generations for natural selection to change a species. a. In long-lived organisms, big changes take millions of years. For example, humans and chimpanzees evolved from a common ancestor that lived 4–12 million years ago. b. Short-lived organisms can evolve rapidly. For example, the flu virus, which has a generation time of about two days, evolves so fast that new vaccines are developed every six months. c. Evolution can be much faster when people decide which individuals survive and reproduce. This is called artificial selection. For example, dogs evolved from wolves and diversified into numerous breeds in less than 40,000 years.

learning objectives encompassing the NGSS DCIs in evolutionary biology (i.e., LS3 and LS4) for grades 3–5 (Table 1). The learning objectives fall into three categories: evidence for evolution (1–3), mechanism of evolution (4–5), and time-scale of evolution (6).

Each group of undergraduates was assigned to a specific elementary school class and introduced to the teacher(s) in week 1. In week 2, the undergraduates traveled to the schools to administer the pre-quiz. Each quiz question and its possible answers were read aloud by one of the undergraduates. After collecting and reviewing the pre-quiz in class, the undergraduates presented and led discussion of an evolutionary topic (e.g., fossils) to further gauge their students' understanding of evolutionary concepts. In weeks 6 and 7, the undergraduates returned to the schools to present their lessons. In week 8, they returned to administer the post-quiz.

The undergraduates were instructed to tailor their lesson plans to address conceptual deficiencies revealed in the pre-quiz and in their discussions with the students, while giving as much weight to learning objective 5 (natural selection) as to any other single learning objective. They were given examples of lesson plans, access to a collection of Pleistocene fossils, and funds for purchasing instructional supplies. They developed and practiced their lesson plans over a four-week period, with multiple rounds of feedback from instructors and peers, and sent two revised drafts to the elementary school teachers for comment. After delivering their first lesson, the undergraduates gave oral reports in which they shared and discussed their classroom experiences with their instructors and peers, with the goal of improving the second lesson plans.

The lesson plans were designed to be taught in two one-hour sessions. Most included a natural selection game and an interactive, phylogeny-building (i.e., evolutionary tree) exercise. Many lessons included fossils, time lines, video clips, and slide shows. Natural selection games were required to include at least two generations to illustrate the response to selection. For lesson plan examples, see Grether et al. (2021).

The undergraduates received instruction in elementary school etiquette and classroom management and were counseled to avoid lecturing, to avoid using unnecessary technical terms, and to define necessary technical terms in a child-friendly way. They were also encouraged to base their lessons on real or at least realistic organisms, not magical creatures or cartoon characters. To prevent them from "teaching to the test," the undergraduates were not allowed to refer back to the pre-quiz questions in their lessons and they were not shown the post-quiz until after they taught their lessons.

Learning Assessment

While the lesson plans varied, the quizzes used to assess learning were the same in all classes. The pre-quiz consisted of six multiple-choice questions, one for each of the six learning objectives, and the post-quiz consisted of six questions of the same type as the pre-quiz followed by two additional questions for learning objectives 3 and 5 (Table 1) that differed structurally from the pre-quiz questions. The purpose of including two different types of questions on the post-quiz was to assess whether the students could generalize what they had learned. Each question had four possible answers, only one of which was correct. Most incorrect answers represented common misconceptions or creationist ideas. The quizzes were written for third-grade comprehension, included pictures, and were printed in color (sample quizzes are included in Grether et al., 2021).

The questions on the pre-quiz in year 1 (2016) served as the first six questions on the post-quiz in year 2 (2018), and the first six questions on the post-quiz in year 1 served as the pre-quiz in

year 2, but the order of the questions and possible answers differed between years. Students were assigned ID numbers for matching their quiz scores within and across years, but no personal identifying information was retained.

This research protocol was reviewed and certified by the UCLA Institutional Review Board (IRB no. 15.001050).

Data Analysis

The response to a quiz question was scored as "correct" if only the correct answer was circled. The sum of correct responses across the first six questions (hereafter "quiz score") was used to compare overall performance between the pre- and post-quiz. I used multilevel mixed-effects general linear regression to analyze quiz scores, and multilevel mixed-effects logistic regression models or Fisher's exact tests to analyze responses to individual quiz questions.

To test for improvement in quiz scores between the pre- and post-quiz across schools, I restricted the analysis to third- and fourth-graders and used multilevel mixed-effects general linear regression with quiz order as the factor and nested random-effects terms for student, class, school, and year (using the mixed command in Stata 14.2). To test for differences between grade levels, I restricted the analysis to school A and used multilevel mixed-effects general linear regression with quiz order, grade level, and their interaction as factors and nested random-effects terms for student, class, and year. The distribution of quiz scores was left-skewed and under-dispersed in relation to a Poisson distribution (i.e., variance < mean). Squaring the quiz scores eliminated the skew and resulted in better Gaussian model fits (as indicated by Wald tests), and those results are presented here, but models with untransformed quiz scores yielded qualitatively similar results (as did multilevel Poisson regression models if they converged). The quiz scores of fifth-graders who had also participated in grade 3 were excluded from these analyses.

To make comparisons between fifth-graders who participated in the study when they were in grade 3 and those who did not, and to compare the third- and fifth-grade quiz scores of students who participated in both years, I used multilevel mixed-effects general linear regression with quiz order and prior participation as factors and nested random-effects terms for student and class. All comparisons were planned, and therefore unadjusted P-values are reported, but the results were qualitatively the same with Sidak adjustments for multiple comparisons.

To test for differences between years in the quiz scores of third- and fourth-graders, I used multilevel mixed-effects general linear regression with quiz order and year as factors and random-effects terms for student, class, and school. To test for variation among classes, I used a mixed-effects general linear regression with quiz order and class as factors and a random-effects term for student.

○ Results

Grade Levels

Across both schools and years, the mean (± SE) quiz score for third- and fourth-graders on the first six quiz questions was 3.50 ± 0.07 on the pre-quiz and 4.54 ± 0.06 on the post-quiz (n = 383 students). Thus, on average, third- and fourth-graders answered one more question correctly on the post-quiz than on the pre-quiz (Figure 1; quiz order effect: χ^2 = 177.96, df = 1, P < 0.0001). Fifth-graders participating in the study for the first time increased from a mean (± SE) of 3.90 ± 0.15 on the pre-quiz to 5.40 ± 0.11 on
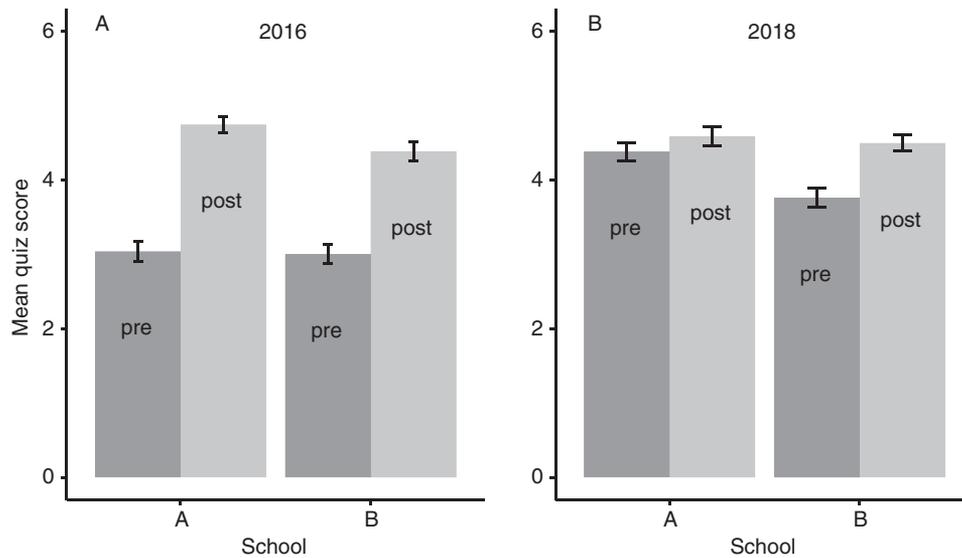
Figure 1. Mean scores (± SE) of third- and fourth-graders on the first six quiz questions, by school and year. "Pre" refers to the pre-quiz (prior to lessons) and "post" refers to the post-quiz (after lessons).

the post-quiz (n = 39 students). Restricting the analysis to school A, where all three grades participated, there was an interaction between quiz order and grade level (Figure 2A; χ^2 = 6.31, df = 2, P = 0.043). Students of all three grade levels scored higher on the post-quiz than on the pre-quiz (grade 3: χ^2 = 37.04, df = 1, P < 0.0001; grade 4: χ^2 = 35.73, df = 1, P < 0.0001; grade 5: χ^2 = 47.56, df = 1, P < 0.0001). There was no significant variation among grade levels in the mean pre-quiz score (χ^2 = 1.77, df = 2, P = 0.42) and no difference between third- and fourth-graders on the post-quiz (χ^2 = 0.63, df = 1, P = 0.43), but fifth-graders scored higher on the post-quiz than the younger students (χ^2 = 20.20, df = 1, P < 0.0001). In terms of improvement in quiz scores between the pre- and post-quiz, there was no difference between third- and fourth-graders (χ^2 = 0.17, df = 1, P = 0.68), while fifth-graders'

scores improved more than those of the younger students (χ^2 = 6.01, df = 1, P = 0.014).

Prior Participation

Among fifth-graders, there was an interaction between quiz order and whether the students had participated in grade 3 (χ^2 = 7.26, df = 1, P = 0.0071). Students who had participated in grade 3 scored higher on the pre-quiz (χ^2 = 7.23, df = 1, P = 0.0072) but not on the post-quiz (χ^2 = 0.64, df = 1, P = 0.42), compared with students who participated for the first time in grade 5 (Figure 2B). Fifth-graders who had participated in grade 3 scored higher on both quizzes than they had in grade 3 (pairwise comparisons; pre-quiz: z = 6.17, P < 0.001; post-quiz: z = 2.54, P = 0.011; n = 31 students), but their
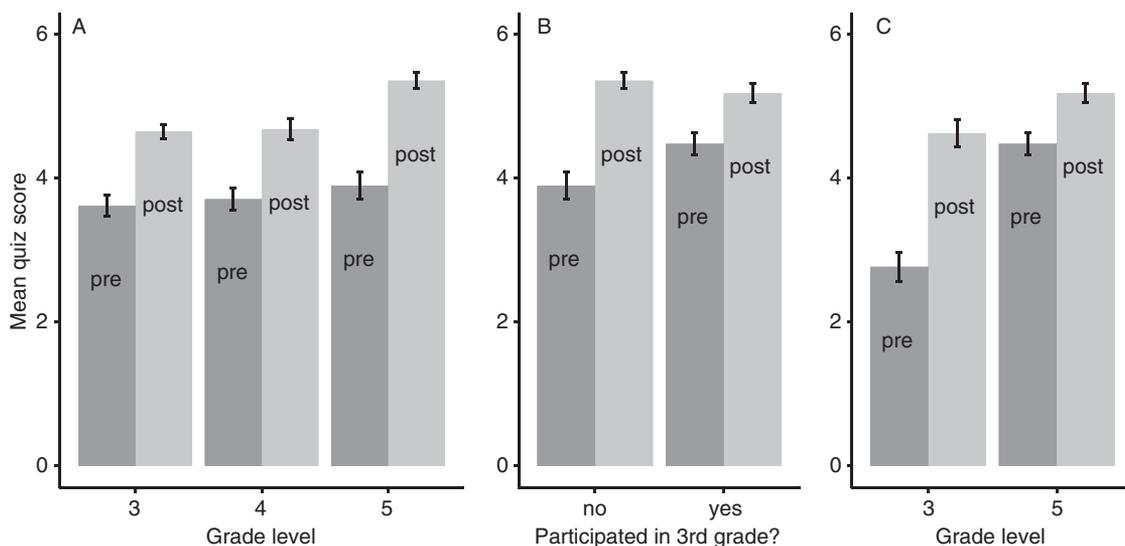

Figure 2. Mean scores (± SE) on the first six quiz questions at school A, **(A)** by grade level, excluding fifth-graders who also participated in grade 3; **(B)** comparing fifth-graders who did or did not participate in grade 3; and **(C)** comparing quiz scores of third- and fifth-grade students who participated in both years.

fifth-grade pre-quiz scores were indistinguishable from their third-grade post-quiz scores ($z = -1.00$, $P = 0.32$; Figure 2C).

Years & Classes

Mean pre-quiz scores of third- and fourth-graders were a full point higher in year 2 (4.02 ± 0.09) than in year 1 (3.01 ± 0.09; $\chi^2 =$ 22.57, df = 1, $P < 0.0001$) but there was no difference between years in post-quiz scores ($\chi^2 = 0.05$, df = 1, $P = 0.83$), resulting in a negative interaction between year and quiz order ($\chi^2 = 44.34$, df = 1, $P < 0.0001$; Figure 1). Thus, there was less improvement in year 2 because the mean pre-quiz score was higher than in year 1.

The third- and fourth-grade classes varied considerably in mean quiz scores ($\chi^2 = 98.91$, df = 1, $P < 0.0001$) and in the degree of improvement between the pre- and post-quiz (quiz order by class interaction; $\chi^2 = 79.18$, df = 17, $P < 0.0001$). Class means ranged from 1.95 to 4.67 on the pre-quiz and from 3.55 to 5.35 on the post-quiz ($n = 18$ classes).

Individual Learning Objectives

Third- and fourth-graders showed improvement on all six learning objectives, although the magnitude of improvement varied (Table 2). The largest improvements were made on the evolutionary time, vestigial traits, and natural selection questions. These students were 3.16 times more likely to answer the natural selection question correctly, 4.18 times more likely to answer the vestigial traits question correctly, and 10.23 times more likely to answer the evolutionary time question correctly on the post-quiz compared to the pre-quiz ($n = 383$ students). They were about twice as likely to correctly answer both types of common ancestry questions and 13.52 times more likely to correctly answer the second natural selection question on the post-quiz, compared to the corresponding pre-quiz questions.

Table 2. Improvement in the quiz scores of third- and fourth-graders between the pre-quiz and post-quiz (N = 383 students). The odds ratio can be interpreted as the increase in the probability of a question being answered correctly on the post-quiz compared to the pre-quiz. For learning objectives 3 and 5, there were two types of questions on the post-quiz, one that was similar to the corresponding pre-quiz question and one that was structurally different.

Learning objective		Odds Ratio	z	P	Question Type
1	Fossils	1.75	2.47	0.013	Same
2	Vestigial traits	4.18	6.16	<0.001	Same
3	Common ancestry	1.86	3.98	<0.001	Same
3	Common ancestry	1.98	4.32	<0.001	Different
4	Heritability	2.96	3.48	<0.001	Same
5	Natural selection	3.16	6.58	<0.001	Same
5	Natural selection	13.52	9.74	<0.001	Different
6	Evolutionary time	10.23	8.31	<0.001	Same

Fifth-graders showed improvement on the questions about vestigial traits (Fisher's exact test, $P = 0.006$), common ancestry ($P < 0.0001$), natural selection ($P = 0.001$), and evolutionary time ($P < 0.0001$), but not on the fossils ($P = 0.12$) and heritability ($P = 0.5$) questions. However, only three fifth-graders answered the fossils question incorrectly and only one answered the heritability question incorrectly on the pre-quiz, and no fifth-graders answered either of these questions incorrectly on the post-quiz ($n = 39$ students).

A majority of students at all grade levels circled the correct answers for the fossils, vestigial traits, and heritability questions on both quizzes (Figure 3). That was not the case for the common ancestry, natural selection, and evolutionary time questions. On the post-quiz, a majority of students circled the correct answers for the natural selection and evolutionary time questions; but, with the exception of fifth-graders, most students still did not circle the correct answer to the common ancestry question.

○ Discussion

The results presented here suggest that two concentrated hours of instruction and active-learning activities can go a long way toward reaching the goals of the NGSS for evolutionary biology in elementary school. Students of all three grade levels showed substantial overall improvement in their understanding of evolutionary concepts, and students who participated in the study in both grades 3 and 5 appeared to retain what they had learned previously. Third- and fourth-graders were more likely to answer every type of question correctly on the post-quiz than on the pre-quiz and showed the most improvement on the vestigial traits, natural selection, and heritability questions (Table 2 and Figure 3). Fifth-graders were also more likely to answer every type of question correctly on the post-quiz than on the pre-quiz and showed the most improvement on the common ancestry, natural selection, and evolutionary time questions. The fossils and heritability questions were the easiest for all grade levels, perhaps because these topics were included in the previous California science standards for grade 2 (California Department of Education, 2004), which were still in effect when these students were in grade 2.

The most challenging concepts for all grade levels were common ancestry and natural selection (Figure 3). Even after the lessons, which invariably emphasized that evolution is a branching process, the concept of an evolutionary ladder, in which "lower" organisms evolve into "higher" organisms, still held sway with a number of elementary school students, as did the idea that some organisms are not related to each other at all. On the natural selection question, the most prevalent misconception, both before and after the lessons, was that changes acquired during an individual's life can be passed on to offspring. Interestingly, all three of these misconceptions align with Lamarck's long-refuted theory of evolution (Mayr, 1972). Very few students in this study thought that fossils or vestigial traits were designed to confuse people. Several students at each grade level selected the creationist "never evolve" answer to the natural selection question on the pre-quiz, but notably fewer students circled this answer on the post-quiz (Figure 3).

The NGSS DCIs for grade 3 include all the evolutionary concepts that the lesson plans in this study were designed to teach. While the results show that third-graders can indeed learn these concepts and retain them at least until grade 5, they also indicate that grade 5 is not too late. Fifth-graders who had participated in the study in grade 3 scored higher on both quizzes than they had
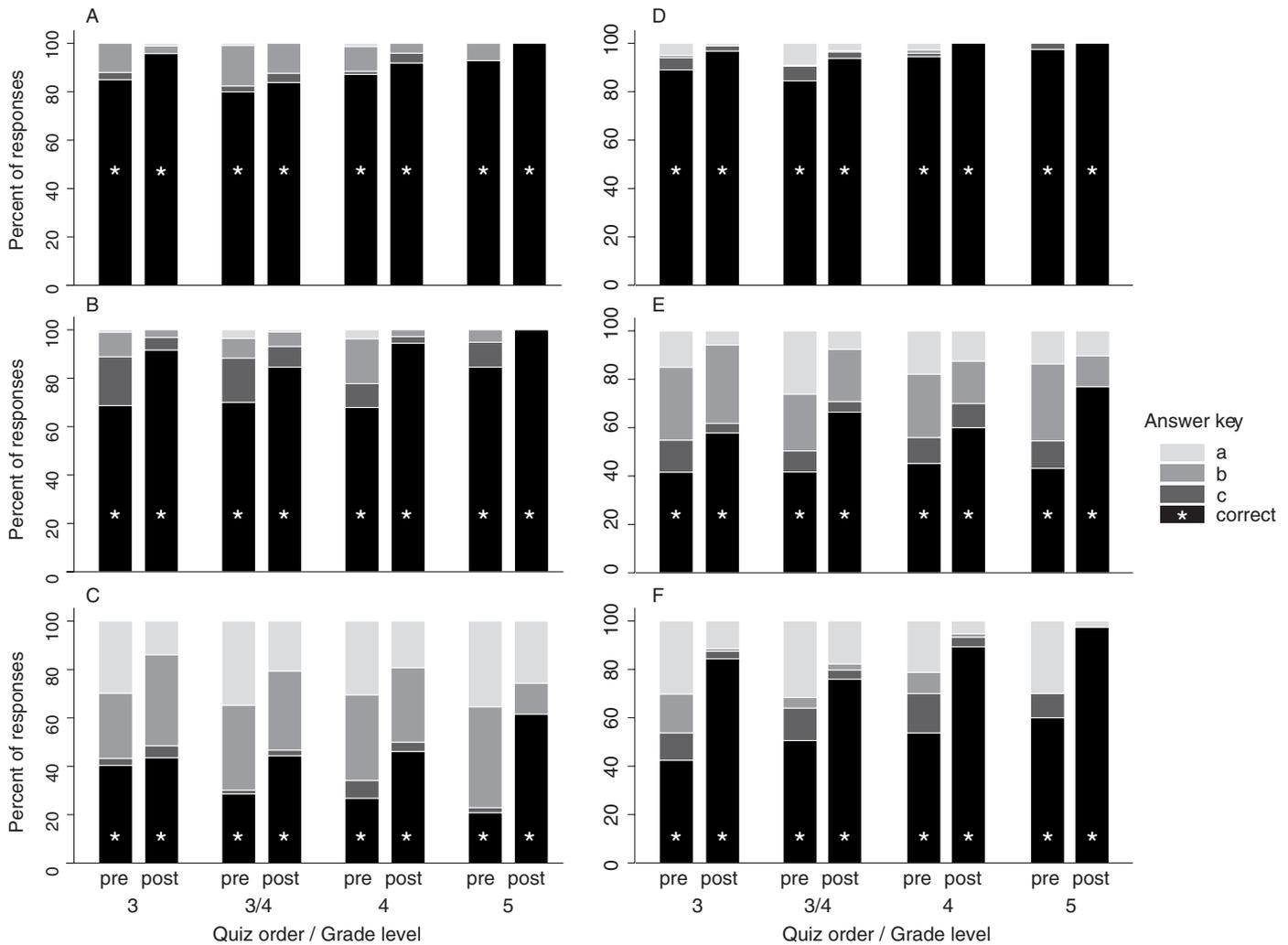
Figure 3. Visual summary of the answers chosen by students before and after the lessons, by quiz question and grade level. Grade levels 3, 4, and 5 correspond to school A, and grade level 3/4 corresponds to school B. Panels A–F represent the first six quiz questions in the same order as the corresponding learning objectives in Table 1. The percent of students answering each question correctly is shown in blue (lowest bar). The other colors (bars) represent different types of wrong answers (see inset answer key). Shorthand descriptions of the wrong answers are as follows: **(A)** Fossils question: (a) to confuse; (b) people put them there; (c) part of the rock. **(B)** Vestigial traits question: (a) to confuse; (b) will evolve trait in the future; (c) had trait earlier in development. **(C)** Common ancestry question: (a) not related; (b) evolutionary ladder; (c) one organism will evolve into the other. **(D)** Heritability question: (a) mystery; (b) want to look like parents; (c) eat same foods as parents. **(E)** Natural selection question: (a) never evolve; (b) acquired characteristics are inherited; (c) individuals change in each generation. **(F)** Evolutionary time question: (a) years; (b) minutes; (c) days. For the actual quiz questions and answers, see Grether et al. (2021).

in grade 3, but their fifth-grade pre-quiz scores were indistinguishable from their third-grade post-quiz scores (Figure 2C). My interpretation is that these fifth-graders retained what they learned in grade 3 but had not advanced in their understanding since then, before the new lessons. However, the finding that fifth-graders participating in the study for the first time achieved post-quiz scores higher than those of third-graders and just as high as those of fifth-graders who had participated previously (Figure 2B) suggests that grade 5 might be a better age to introduce the most challenging concepts. By contrast, there was no indication that fourth-graders were better at mastering these concepts than third-graders (Figures 2A and 3). From the standpoint of teaching evolutionary concepts, it would be ideal to repeat them at all grade levels, but classroom

time is limited and teachers have other science standards to meet. Therefore, if these results are substantiated by further research, an adjustment to the NGSS guidelines might be warranted. Spacing the evolutionary biology DCIs out, as the NRC originally proposed (National Research Council, 2012), might be a better strategy than concentrating them all in grade 3.

Students could potentially learn how to answer particular types of questions without actually learning the underlying concepts. To address this issue, I included two different types of questions about common ancestry and natural selection on the post-quiz. One of the two questions was directly analogous to the corresponding pre-quiz question while the other question was of a new structure, with different types of incorrect answers. The results indicate that the

students were able to generalize what they learned from one type of question to another (Table 2).

Because there was no replication of lesson plans, it is impossible to draw firm conclusions about which lesson plans were most effective. The student composition and regular classroom teachers undoubtedly account for much the variation among class means. Another possible shortcoming of this study is that the pre-quiz included only six multiple-choice questions, one per learning objective. Including more questions, of varied types, and using other methods of assessment, such as interviewing students individually before and after the lessons, would have provided greater resolution of the students' grasp of evolutionary concepts. However, in studies of this sort, the possible dividends of asking a larger number of questions and using other methods of assessment need to be balanced against the constraints of available classroom time and the attention spans of children.

○ Acknowledgments

References

Anderson, C.W., de los Santos, E.X., Bodbyl, S., Covitt, B.A., Edwards, K.D., Hancock, J.B., et al. (2018). Designing educational systems to support enactment of the Next Generation Science Standards. *Journal of Research in Science Teaching, 55,* 1026–1052.

Baum, D.A., Smith, S.D. & Donovan, S.S.S. (2005). The tree-thinking challenge. *Science, 310,* 979–980.

Binder, W.J. & Van Valkenburgh, B. (2010). A comparison of tooth wear and breakage in Rancho La Brea sabertooth cats and dire wolves across time. *Journal of Vertebrate Paleontology, 30,* 255–261.

California Department of Education (2004). *Science Framework for California Public Schools Kindergarten through Grade Twelve with New Criteria for Instructional Materials.* Sacramento, CA: California Department of Education.

California Department of Education (2019). *NGSS for California Public Schools, K–12.* Retrieved from https://www.cde.ca.gov/pd/ca/sc/ngssstandards.asp.

Campos, R. & Sá-Pinto, A. (2013). Early evolution of evolutionary thinking: teaching biological evolution in elementary schools. *Evolution: Education and Outreach, 6,* article 25.

Carbone, C., Maddox, T., Funston, P.J., Mills, M.G.L., Grether, G.F. & Van Valkenburgh, B. (2009). Parallels between playbacks and Pleistocene tar seeps suggest sociality in an extinct sabretooth cat, *Smilodon. Biology Letters, 5,* 81–85.

Clark, R.E., Kirschner, P.A. & Sweller, J. (2012). Putting students on the path to learning: the case for fully guided instruction. *American Educator* (Spring), 6–11.

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

Dorph, R., Goldstein, D., Lee, S., Lepori, K., Schneider, S. & Venkatesan, S. (2007). The status of science education in the Bay Area: research brief. Lawrence Hall of Science, University of California, Berkeley. Retrieved from http://static.lawrencehallofscience.org/rea/bayareastudy/.

Emmons, N.A. & Kelemen, D.A. (2015). Young children's acceptance of within-species variation: implications for essentialism and teaching evolution. *Journal of Experimental Child Psychology, 139,* 148–160.

Emmons, N., Lees, K. & Kelemen, D. (2018). Young children's near and far transfer of the basic theory of natural selection: an analogical storybook intervention. *Journal of Research in Science Teaching, 55,* 321–347.

Evans, E.M. (2000). The emergence of beliefs about the origins of species in school-age children. *Merrill-Palmer Quarterly, 46,* 221–254.

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

Grether, G.F. (2010a). Evolution. In D. Mills (Ed.), *The Encyclopedia of Applied Animal Behaviour & Welfare.* Cambridge, MA: CABI.

Grether, G.F. (2010b). Selection. In D. Mills (Ed.), *The Encyclopedia of Applied Animal Behaviour & Welfare.* Cambridge, MA: CABI.

Grether, G.F., Chock, R.Y., Cowen, M.C., De La Cruz-Sevilla, J.S., Drake, T.N., Lum, K.S., et al. (2021). Teaching evolutionary concepts in elementary school. *American Biology Teacher, 83,* xxx–xxx.

Hermann, R. S. (2011). Breaking the cycle of continued evolution education controversy: on the need to strengthen elementary level teaching of evolution. *Evolution: Education and Outreach, 4,* 267–274.

Hmelo-Silver, C.E., Duncan, R.G. & Chinn, C.A. (2007). Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner. *Educational Psychologist, 42,* 99–107.

Kirschner, P.A., Sweller, J., Clark, R.E., Kirschner, P.A., Sweller, J. & Clark, R.E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41,* 75–86.

Krajcik, J., Codere, S., Dahsah, C., Bayer, R. & Mun, K. (2014). Planning instruction to meet the intent of the Next Generation Science Standards. *Journal of Science Teacher Education, 25,* 157–175.

Lucci, K. & Cooper, R.A. (2019). Using the I2 strategy to help students think like biologists about natural selection. *American Biology Teacher, 81,* 88–95.

Mayr, E. (1972). Lamarck revisited. *Journal of the History of Biology, 5,* 55–94.

Mervis, J. (2015). Why many U.S. biology teachers are 'wishy-washy.' *Science, 347,* 1054.

Nadelson, L., Culp, R., Bunn, S., Burkhart, R., Shetlar, R., Nixon, K. & Waldron, J. (2009). Teaching evolution concepts to early elementary school students. *Evolution: Education and Outreach, 2,* 458–473.

National Research Council (2012). Disciplinary core ideas – life sciences. In *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (pp. 139–168). Washington, DC: National Academies Press.

National Science Teachers Association (2019). NGSS@NSTA: STEM starts here. Retrieved from https://ngss.nsta.org.

NGSS Lead States (2013). *Next Generation Science Standards: For States, by States.* Washington, DC: National Academies Press.

Padian, K. (2013). Correcting some common misrepresentations of evolution in textbooks and the media. *Evolution: Education and Outreach, 6,* article 11.

Prinou, L., Halkia, L. & Skordoulis, C. (2011). The inability of primary school to introduce children to the theory of biological evolution. *Evolution: Education and Outreach, 4,* 275–285.

Ripple, W.J. & Van Valkenburgh, B. (2010). Linking top-down forces to the Pleistocene megafaunal extinctions. *BioScience, 60,* 516–526.

Rosenshine, B. (2012). Principles of instruction: research-based strategies that all teachers should know. *American Educator* (Spring), 12–20.

Watanabe, T. (2011). California teachers lack the resources and time to teach science. *Los Angeles Times*, October 31.

Young, A.K., White, B.T. & Skurtu, T. (2013). Teaching undergraduate students to draw phylogenetic trees: performance measures and partial successes. *Evolution: Education and Outreach, 6*(6), 1–15.

GREGORY F. GRETHER is a Professor in the Department of Ecology & Evolutionary Biology, University of California, Los Angeles, CA 90095; e-mail: ggrether@ucla.edu.

Affiliate Members

Biology Teachers Association of New Jersey (BTANJ)
Colorado Biology Teachers Association (CBTA)
Cleveland Regional Association of Biologists (CRABS)
Connecticut Association of Biology Teachers (CTABT)
Delaware Association of Biology Teachers (DABT)
Empire State Association of Two-Year College Biologists (ESATYCB)
Hong Kong Association of Biology Teachers (HKABT)
Illinois Association of Biology Teachers (IABT)
Illinois Association of Community College Biologists (IACCB)
Indiana Association of Biology Teachers (IABT)

Kansas Association of Biology Teachers (KABT)
Louisiana Association of Biology Teachers (LABT)
Massachusetts Association of Biology Teachers (MABT)
Michigan Association of Biology Teachers (MABT)
Mississippi Association of Biology Educators (MSABE)
Missouri Association of Biology Teachers (MOBioTA)
New York Biology Teachers Association (NYBTA)
South Carolina Association of Biology Teachers (SCABT)
Tennessee Association of Biology Teachers (TNABT)
Texas Association of Biology Teachers (TABT)
Virginia Association of Biology Teachers (VABT)

The National Association of Biology Teachers supports these affiliate organizations in their efforts to further biology & life science education.