Fostering 21st-Century Evolutionary Reasoning: Teaching Tree Thinking to Introductory Biology Students

Laura R. Novick[†]* **and Kefyn M. Catley**[‡]

[†]Department of Psychology and Human Development, Peabody College, Vanderbilt University, Nashville, TN 37203-5721; [‡]Department of Biology, Western Carolina University, Cullowhee, NC 28723

ABSTRACT

The ability to interpret and reason from Tree of Life (ToL) diagrams has become a vital component of science literacy in the 21st century. This article reports on the effectiveness of a research-based curriculum, including an instructional booklet, laboratory, and lectures, to teach the fundamentals of such tree thinking in an introductory biology class for science majors. We present the results of a study involving 117 undergraduates who received either our new research-based tree-thinking curriculum or business-as-usual instruction. We found greater gains in tree-thinking abilities for the experimental instruction group than for the business-as-usual group, as measured by performance on our novel assessment instrument. This was a medium size effect. These gains were observed on an unannounced test that was administered ~5–6 weeks after the primary instruction in tree thinking. The nature of students' postinstruction difficulties with tree thinking suggests that the critical underlying concept for acquiring expert-level competence in this area is understanding that any specific phylogenetic tree is a subset of the complete, unimaginably large ToL.

Scientists' ability to interpret and reason with information depicted in the Tree of Life (ToL)—that is, to engage in a suite of skills referred to as tree thinking—has yielded important benefits for humanity in many areas, including agriculture, biotechnology, climate change, forensics, and health (e.g., American Museum of Natural History [AMNH], 2002; Futuyma, 2004; Thomas *et al.*, 2004; Yates *et al.*, 2004; Davis *et al.*, 2010). Thus, the ability to engage in tree thinking is an important component of 21st-century science literacy (Thanukos, 2009; Baum and Smith, 2013; Novick and Catley, 2013). In recognition of this, tree thinking has recently become a rich area of empirical inquiry. This research has primarily examined undergraduates' ability to interpret branching tree diagrams called cladograms, which depict (hypothesized) phylogenetic relationships as nested sets of taxa supported by synapomorphies (Hennig, 1966; Thanukos, 2009; Baum and Smith, 2013). For example, the cladogram in Figure 1 depicts evolutionary relationships among 10 dinosaur taxa. Research indicates that undergraduates have difficulty engaging in tree thinking (e.g., Meir *et al.*, 2007; Novick and Catley, 2013), even after instruction in phylogenetics in a college biology class (Sandvik, 2008; Halverson *et al.*, 2011; Catley *et al.*, 2012; Phillips *et al.*, 2012; Dees *et al.*, 2014). Moreover, understanding natural selection and success at tree thinking are distinct constructs, so current instruction that focuses primarily on topics in microevolution (Catley, 2006) is insufficient for promoting competence at tree thinking, which is a macroevolutionary skill (Novick *et al.*, 2014).

Clearly, there is a need for improved instruction in tree thinking that 1) is informed by a deep understanding of both the relevant biological science and the difficulties students encounter and 2) leverages knowledge of effective instructional practices (National Research Council, 2012). Several research teams have recently begun to

Jennifer Momsen, *Monitoring Editor*
Submitted June 8, 2015; Revised May 3, 2016; Accepted May 9, 2016
CBE Life Sci Educ December 1, 2016 15:ar66
DOI:10.1187/cbe.15-06-0127

*Address correspondence to: Laura R. Novick, (Laura.Novick@vanderbilt.edu).

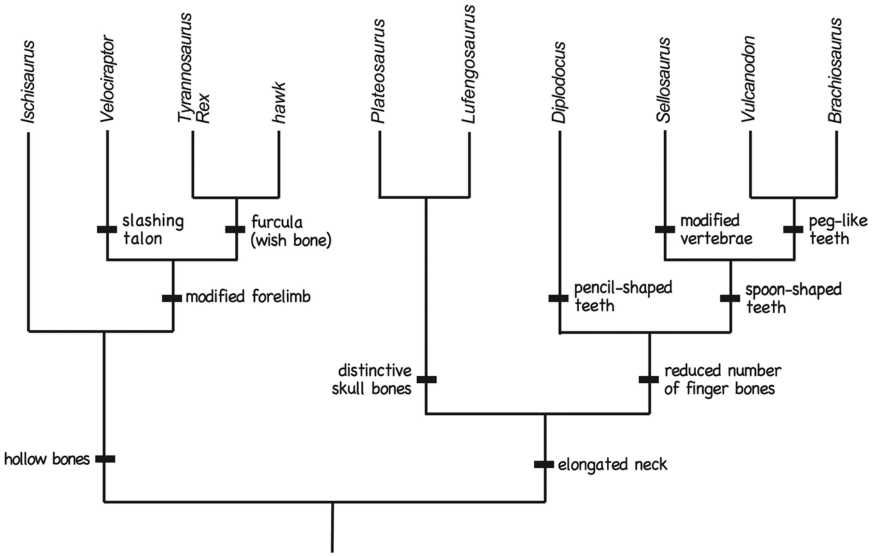

FIGURE 1. A cladogram that appeared on the tree-thinking assessments. Adapted with permission of Springer Science+Business Media from Figure 3 in Catley *et al.* (2013, p. 2334). (Modifications were to [a] rotate the four branches at one node, [b] print the taxon names in italics, and [c] change the character markers from circles to horizontal lines.)

develop tree-thinking curricula (Singer *et al.*, 2001; Smith and Cheruvelil, 2009; Smith *et al.*, 2013; Eddy *et al.*, 2013; McLaurin *et al.*, 2013; Novick *et al.*, 2014), although the extent to which these efforts were informed by empirical research on students' difficulties varies. The primary goal of the present study was to evaluate the feasibility and effectiveness of moving our research-based tree-thinking instruction (Novick *et al.*, 2014) from the highly controlled setting in which it was validated to an introductory biology class for science majors. We also report on our development and testing of a new phylogenetics laboratory that is appropriate for introductory biology classes or as part of an extended tree-thinking curriculum that could be included in a variety of organismal biology classes.

DESIGNING A RESEARCH-BASED TREE-THINKING CURRICULUM

What Students Need to Know about Phylogenetic Trees

Competence in the area of phylogenetics involves being able to read trees, make inferences, and build the nested structure of trees from character data. Although a basic knowledge of how trees are built is helpful for understanding how to interpret and reason from trees, in general, tree-building is a more advanced topic. Thus, our instruction focused on how to interpret trees, what we referred to earlier as tree thinking. Eleven tree-thinking skills are listed in Table 1, along with supporting references. The seven skills labeled parenthetically as I–VII were collectively described by Novick and Catley (2013).

Students should be able to 1) identify a character (technically, a synapomorphy) that two or more taxa share due to inheritance from their most recent common ancestor (MRCA) and, 2) correspondingly, identify a set of taxa that share a certain character. We will use the scientific term "synapomorphy" and the less precise but more familiar term "character" interchangeably.

Students also must understand the relational/structural information depicted in cladograms. In part, this means being able to determine whether a set of taxa comprise a clade and being able to identify all the nested clades. Clades include an MRCA and all its descendants and are the only valid biological groups.

Another critical relational skill is the ability to assess the relative evolutionary relatedness of taxa, given both resolved and polytomous structures. This is perhaps the most important tree-thinking skill, as the purpose of cladograms is to depict evolutionary relationships. The basic unit of cladogram structure is a three-taxon statement in which two taxa in a set of three share a more recent common ancestor with each other than they do with the third taxon. This structure describes a set of relationships that are resolved. In the cladogram in Figure 1, *Velociraptor*, *Tyrannosaurus rex*, and hawk, for example, comprise a three-taxon statement (as do hawk, *Diplodocus*, and *Vulcanodon*). In contrast, in the cladogram in Figure 2, rabbit, mole, and raccoon comprise a polytomy (as do rabbit, mole, and either skunk or dog): No two of these three taxa share a more recent common ancestor with each other than with the third taxon in the group.

Inference from phylogenetic relationships is critical in both basic and applied biology and thus constitutes another tree-thinking skill. For example, scientists have used phylogenetic evidence summarized in cladograms to infer 1) soft-tissue morphological characters of extinct taxa known only from fossils (e.g., Bryant and Russell, 1992; Witmer, 1995) and 2) the appropriate antivenin to use to treat the bite of a snake whose venin is unknown (AMNH, 2002).

Yet another skill involves identifying the sequential order of appearance of characters on a given evolutionary path. For example, the sequence of characters elongated neck, reduced number of finger bones, and spoon-shaped teeth provides evidence for the evolutionary relationship between *Sellosaurus* and *Vulcanodon* (see Figure 1). Conversely, cladograms also provide evidence for convergent evolution, through the appearance of a similar character on multiple branches.

Another important skill, subsets of the ToL, involves being able to reason about common relationships given changing subsets of taxa across multiple trees. Students need to understand that adding taxa to or removing taxa from a particular tree does not change the relationships among the taxa that are in all the trees. For example, in Figure 2, the relationship among rabbit, mole, and raccoon remains the same if skunk and dog are removed or if hedgehog is added as the sister group to mole. Finally, students need to understand that rotating cladogram branches around their nodes does not change the relationships among the affected taxa, even though the adjacency relations among the taxa do change.

Summary of Our Previous Curriculum Work

The development of the tree-thinking instruction used in the present study follows directly from our earlier work. Although

TABLE 1. Eleven tree-thinking skills

Skill name[a]	Description	References
Identify characters (I)	Identify a synapomorphy that two or more taxa share due to inheritance from their MRCA.	Meir *et al.* (2007); Novick and Catley (2013)
Identify taxa (II)	Identify a set of taxa that share a certain character.	Novick and Catley (2013)
Identify/evaluate clades (III)	Evaluate whether a given set of taxa comprises a clade.	Hennig (1966); Thanukos (2009); Baum and Smith (2013); Novick and Catley (2013)
Identify nested clades	Mark all the nested clades in a cladogram.	Meisel (2010); Baum and Smith (2013); Novick *et al.* (2014)
Evolutionary relationship: resolved structure (IV)	Assess relative evolutionary relatedness when three taxa are resolved (i.e., comprise a three-taxon statement).	Baum *et al.* (2005); Baum and Smith (2013); Novick and Catley (2013)
Evolutionary relationship: polytomy	Assess relative evolutionary relatedness when three taxa comprise a polytomy (i.e., no two of the taxa share a more recent common ancestor with each other than with the third taxon).	Baum *et al.* (2005); Baum and Smith (2013); Novick *et al.* (2014)
Inference (V)	Use the information depicted in a cladogram to make an inference based on phylogenetic relationship.	Novick and Catley (2013)
Evolutionary sequence (VI)	Identify the sequential order of appearance of characters on a designated evolutionary path.	Novick and Catley (2013)
Convergent evolution (VII)	Recognize that characters that appear on multiple branches of a cladogram are indicative of convergent evolution.	Novick and Catley (2013)
Subsets of the ToL	Reason about common relationships in the face of changing subsets of taxa.	Baum *et al.* (2005); Baum and Smith (2013); Novick *et al.* (2014)
Rotation	Rotating cladogram branches around their nodes does not change the relationships among the affected taxa, even though the adjacency relations among the taxa do change.	Baum *et al.* (2005); Gregory (2008); Baum and Smith (2013)

[a]Skills numbered I–VII were collectively described by Novick and Catley (2013).

some of what we implemented in the current study is similar to recent work by other researchers, the development work proceeded independently and in parallel. We will note these similarities as we discuss the components of our curriculum.

Our initial curriculum design work involved creating a short, self-paced instructional booklet to teach tree thinking to college students (Novick *et al.*, 2014). We also developed an assessment, based largely on items validated in prior experimental work, to assess students' ability to employ several key tree-thinking skills and to evaluate the efficacy of our instruction. The results of a large-scale study involving college students with both weaker and stronger backgrounds in biology demonstrated the efficacy of this instructional booklet. Students who were randomly assigned to receive the instructional booklet, which took ~30 minutes to complete, did much better on the tree-thinking assessment than students who were randomly assigned to the control (no instruction) condition. With a Cohen's d of 1.47 for a composite measure of tree thinking, the benefit of the short instructional booklet was impressive. In a comparison of the stronger background students who received the instruction to similar students in earlier studies who had received 2 days of instruction in phylogenetics in their college biology class, we found that our short booklet was as effective as the classroom instruction for teaching students to reason about clades and more effective at conveying an understanding of evolutionary relatedness, especially for taxa in polytomous relationships.

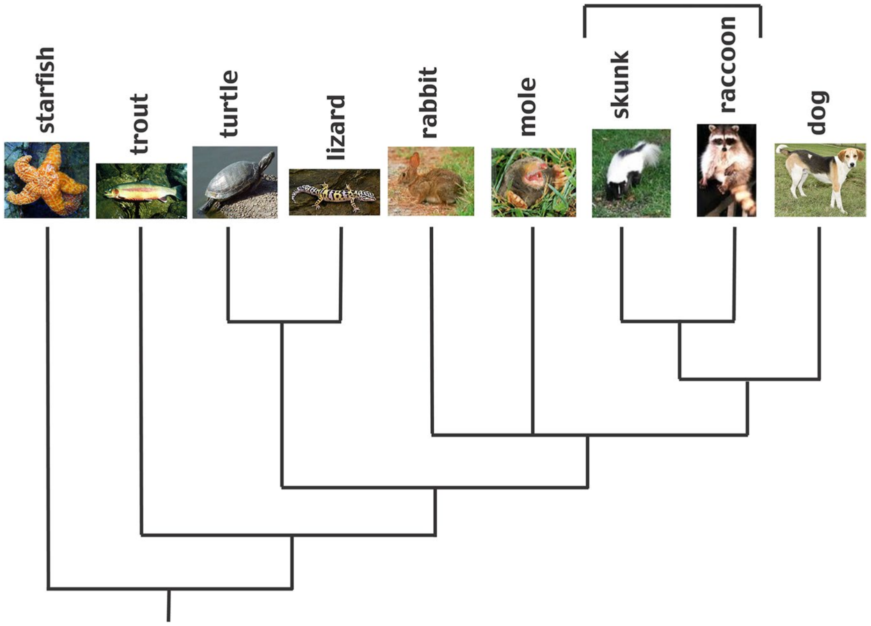

FIGURE 2. A cladogram that appeared on the tree-thinking assessments. Students received a version of this cladogram that included color photographs. Adapted with permission of Springer Science+Business Media from Figure 1 in Phillips *et al.* (2012, p. 596). (The modification was to include the original color photograph in the online publication.)

This success is laudable. Nevertheless, this study has three limitations, which concern the setting in which the research was implemented, the effectiveness of the instructional booklet, and the design of the assessment. First, and most importantly, because the instruction was administered in a controlled setting with an immediate posttest, it is unknown whether it would be effective in the more naturalistic and less well controlled, classroom environment, especially with a delayed posttest that is typical for college classes. Second, although students did quite well on the clade and inference test items, there was room for improvement in their ability to evaluate evolutionary relationships in both resolved and polytomous topologies. Other studies confirm that evolutionary relatedness is a particularly difficult concept for students to understand (e.g., Dees *et al.*, 2014). Third, most of the items intended to assess students' understanding that cladograms are subsets of the larger ToL turned out to be poorly conceived, as they could be answered using general knowledge of hierarchies rather than requiring an understanding of concepts specific to reasoning about evolutionary trees.

Overview of the Present Study

The present study addressed all three of these limitations: using students enrolled in a large undergraduate biology class for science majors, we compared experimental versus business-as-usual instruction in tree thinking using a pretest/posttest design. Students in K.M.C.'s two sections received our research-based tree-thinking instruction; those in the other two sections received business-as-usual instruction. All students completed a pretest before instruction and an identical posttest at the end of the course. The instructional booklet developed by Novick *et al.* (2014) was edited to improve the clarity of the section on evaluating evolutionary relationships. We wrote new test items to assess students' understanding of the subsets of the ToL skill.

We also developed a new component of our tree-thinking curriculum for this study, a laboratory on phylogenetics. The biology faculty thought the phylogenetics laboratory then being used in the introductory biology class was suboptimal. In return for being allowed to test all students for our study, we agreed to write a new research-based phylogenetics laboratory that would provide better coverage of the topic. Thus, all students received part of our experimental tree-thinking curriculum.

Because our new tree-thinking curriculum was a normal part of the instructional content of the introductory biology class, our study was designated as "exempt" by the Western Carolina University Institutional Review Board and consent forms were not required.

METHODS

Students

Students (N = 135) enrolled in the second-semester introductory biology course for science majors at Western Carolina University participated in this study—66 in the sections that received business-as-usual instruction and 69 in the sections that received our research-based tree-thinking instruction. This course presents an introduction to the major eukaryote unicellular, plant, fungi, and animal taxa. The data from 18 students who took only one test (14 pretest, four posttest) were excluded from the analyses. We obtained complete data from 55 and 62 students in the business-as-usual and experimental conditions, respectively. Table 2 documents the similarity of the students in the two conditions based on the demographic information available. Although the distribution of students' majors differed somewhat across the two conditions, the difference was not statistically significant. The smaller number of biology majors in the experimental condition was offset by a larger number of other science majors. The most common such majors (in both conditions) were forensics and natural resource management. Given that one might expect biology majors to have an advantage on tests of explicitly biological concepts, to the extent that students' majors differ across conditions, the difference works against our hypothesis of finding larger improvements in tree thinking in the experimental instruction condition.

TABLE 2. Characteristics of the students in the business-as-usual and experimental instruction conditions and the phylogenetics instruction they received before the onset of the instructional manipulation

	Tree-thinking instructional condition	
	Business-as-usual	**Experimental**
Sample size (for analysis)	55	62
Mean year in school[a]	2.35	2.61
Females[b]	47%	60%
Biology majors[c]	47%	26%
Other science majors[c]	36%	52%
Science education majors[c]	7%	5%
All other majors[c]	9%	18%
Class lectures	3 days/week for 50 minutes each	2 days/week for 75 minutes each
Phylogenetics content, first half of the class (first instructor)	1) One lecture 2) Textbook chapter 26 ("Phylogeny and the Tree of Life"), which includes 16 cladograms 3) Trees were shown in later lectures to orient students to each new group being considered, but neither the trees nor tree thinking were the focus of discussion.	1) One lecture 2) Part of textbook chapter 26 was assigned 3) Trees were shown in later lectures to orient students to each new group being considered, but neither the trees nor tree thinking were the focus of discussion.
Phylogenetics lab	Yes	Yes

[a] Sophomore = 2; junior = 3. $F(1, 115) = 2.47$, $p > 0.10$, MSE = 0.85, $\eta_p^2 = 0.02$.

[b] $\chi^2(1, N = 117) = 1.81$, $p > 0.15$.

[c] $\chi^2(3, N = 117) = 7.15$, $p > 0.06$.

Design

Experimental versus business-as-usual instruction in phylogenetics varied between subjects, with students assigned to conditions based on the class section in which they were enrolled. Time of test was manipulated within subjects, as all students took the tree-thinking assessment both before and after instruction. The tests were given during the seven laboratory sections of the course (taught by two graduate students), which met once a week and included students from all four lecture sections.

Course Organization

In each lecture section, one professor taught the first half of the course, which covered unicellular and plant taxa, and another professor taught the second half of the course, which covered fungi and animal taxa. One team taught the two business-as-usual sections; a different team taught the two experimental instruction sections. During the first half of the course, students in the two conditions had similar instruction in phylogenetics (see Table 2). The experimental instruction occurred during the second half of the course. All sections of the course used Campbell *et al.*'s (2008) textbook, which has 66 cladograms, primarily in the chapters on animal and plant diversity.

Instructional Conditions

The professor who taught the second half of the business-as-usual sections spent ~40 minutes discussing phylogenetics and most recent common ancestry. A few exam questions asked about which taxa are more closely related to which other taxa.

A brief summary of our experimental instruction is provided here. More details can be found in Section A of the Supplemental Material. Students received a personal copy of Novick *et al.*'s (2014) instructional booklet (slightly revised), which was described as a supplement to chapter 26 in the textbook, which was also assigned. They were told to read the booklet and complete the two practice-what-you-learned sections for homework. These students then received 2.5 hours (2 days) of lecture covering phylogenetic concepts. The slides provided with the textbook were used, but they were upgraded and annotated. Following these introductory lectures, a phylogenetics perspective was adopted for the rest of the course material by introducing each new taxon in terms of the synapomorphies that provide evidence for its phylogenetic placement, as illustrated in a cladogram. The instructional booklet and subsequent lectures reinforced and extended many of the concepts covered in the phylogenetics laboratory, thus helping to provide a unified learning experience across the two parts of the course (see Smith and Cheruvelil, 2009). Throughout the remainder of the semester, students were encouraged to use tree thinking as a powerful tool for learning, organizing, and retrieving information.

Procedure

The pretest was given the week before the phylogenetics laboratory and took ~30 minutes to complete. Most laboratories had a prelab assignment that counted toward students' lab grades. Our pretest was the prelab assignment for the phylogenetics laboratory. Students were told they would not know the answers to all the questions because they probably had not learned much of the relevant material yet. However, if they took the assignment seriously and did their best, they would receive full credit; otherwise, they would receive no credit. All students appeared to take the assignment seriously, so all received full credit. The laboratory was completed the following week under the direction of the laboratory instructors. The experimental instruction began during the next class period.

The posttest, which was worth 25% of students' laboratory grades, was given 7 weeks after the pretest, during the last laboratory class.[1] It was identical to the pretest. The posttest was unannounced, because we felt uncomfortable telling students in the business-as-usual condition to study for a test on concepts that were introduced in a laboratory completed nearly 2 months earlier and that had not been covered extensively in their lecture class since that time.

Phylogenetics Laboratory

A brief summary of the new phylogenetics laboratory is provided here. The laboratory materials included a student laboratory manual, an instructors' guide (with answer key), and various specimens. More details about the laboratory can be found in Section B of the Supplemental Material. The laboratory took students most of the 3-hour class period to complete.

The student laboratory manual was an 11-page, six-part booklet that began by describing phylogenetics as the study of the history of life. Part I asked students, who worked in small groups, to examine members of nine major groups of animal taxa (Annelida, Arthropoda, Chordata, Cnidaria, Echinodermata, Mollusca, Nematoda, Platyhelminthes, and Porifera) to determine how the different possible states of each of 11 characters (nine synapomorphies, two convergently evolved characters) are distributed among the groups.

Part II asked students to map these character states onto three alternative cladogram topologies, which were introduced as hypotheses of the relationships among the nine animal groups. In part III, students evaluated these topologies to determine which provides the best representation of the historical evolutionary relationships among the taxa by considering two criteria, which we explained: parsimony and how much of the topology is resolved. In discussing the latter criterion, we explained the difference between resolved relationships and polytomies.

Part IV taught students to recognize the difference between homologies and homoplasies—that is, characters that are shared by taxa due to shared ancestry versus to independent (convergent) evolution. In part V, students were told that one reason why cladograms are useful is that they provide a powerful basis for making inferences. An example that included a simple inference problem was given. Part VI presented four extension questions for class discussion.

Instructional Booklet

The instructional booklet was a slightly revised version of that developed and validated by Novick *et al.* (2014). It is self-paced and takes ~30 minutes to complete. The pedagogical features of the booklet are described in Novick *et al.* (2014). Briefly, the booklet began with foundational terminology and

[1] The students who received our research-based tree-thinking instruction had significantly higher scores on the posttest than did students who received business-as-usual instruction in phylogenetics. Because it would not be fair to penalize students with respect to their laboratory course grades because they happened to enroll in one of the latter sections of the lectures, we converted the raw posttest scores for students in the business-as-usual condition to scale scores that had the same mean and SD as the raw scores of students in the experimental instruction condition.

concepts, including taxon, synapomorphy, cladogram, MRCA, clade, and three-taxon statement. Consistent with Meisel's (2010) recommendation, the nested hierarchical structure of cladograms was stressed. This was followed by an in-depth discussion of how to determine relative evolutionary relatedness among taxa by determining which taxa share a more recent common ancestor. To more clearly distinguish the scientifically appropriate method from other, inappropriate methods, we revised this section by dividing it into named subsections for most recent common ancestry and for each of two misconceptions about how to determine evolutionary relatedness: horizontal distance between taxa and number of vertical steps between taxa. The section explaining the concept of a polytomy was also revised in light of the difficulties Novick *et al.*'s (2014) students had on those test items. The final section discussed cladograms as subsets of the complete ToL. In this section, students were taught how to prune and collapse taxa, thereby creating a cladogram with fewer branches, and how to merge separate smaller cladograms into a single, larger cladogram. More information about the instructional booklet can be found in Section A of the Supplemental Material.

Tree-Thinking Assessment

We began with Novick *et al.*'s (2014) validated assessment. A variety of types of evidence can be used to support the validity of an assessment for the intended interpretation of the scores (American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education, 1999; Kaplan and Saccuzzo, 2009). With respect to content-related validity, the tree-thinking skills were identified by K.M.C. (also see Novick and Catley, 2013) based on his doctoral training in systematics, professional expertise (e.g., Catley, 1994), and extensive experience teaching this area of biology (including a semester-long evolution course). Moreover, the specific questions call for the kinds of reasoning in which professional biologists engage when examining phylogenetic trees (e.g., see Baum and Smith, 2013). Evidence that scores on the assessment are related to performance on an external criterion provides support for the criterion-related validity of the assessment. As we described in Novick *et al.* (2014), for the questions that were taken from our extensive research on college students' ability to engage in tree thinking, there is a close relationship between tree-thinking skill and having taken more biology classes in which topics related to macroevolution were likely to have been covered.

We modified Novick *et al.*'s (2014) assessment where necessary to replace ineffective questions and assess new skills. Also, to simplify scoring the explanation questions for classroom use, we changed those items from free response to multiple choice. Twelve explanations were provided (see Table 3): 11 involved (appropriate and inappropriate) concepts used by students in their free responses to these questions on our earlier assessment (Novick *et al.*, 2014); the last was none of the above. The explanations were introduced in a box on the instruction page of the test booklet and were reprinted on each test page that included an explanation question. Students selected the explanation that best fit their reasoning.[2]

TABLE 3. The explanation choices provided to students on the assessments

Code	Explanation[a]
CATEG	These taxa are in the same category (e.g., bugs, mammals). *Name the category.*
CHAR	These taxa have more characteristics in common with each other.
CLOSE	These taxa are close(r) together, looking left to right across the diagram.
COM_A	These taxa have a common ancestor. *Mark on the diagram.*
CONN	These taxa are connected (or come from the same line) on the diagram.
CONV	This is an example of convergent evolution.
DESC	Not all the descendants of the most recent common ancestor are included.
INTO	One taxon on the cladogram evolved into (i.e., is the ancestor of) the other(s). *Mark the original taxon on the diagram.*
M_REC	These taxa share a more/most recent common ancestor. *Mark on the diagram.*
STEPS	There are fewer steps between these taxa, looking at the branching up and down in the diagram.
TAXA	These are all the taxa that share a most recent common ancestor.
NONE	None of these explanations is close to the reason why I gave the answer I did.

[a]The four directives printed in italics here (e.g., "*Name the category*" for the CATEG explanation) were printed in red ink in the test booklets to draw attention to them. However, students almost universally ignored them. Thus, they will not be discussed further. The central idea of each explanation was underlined to help students make sense of the abbreviations.

Table 4 lists the types of questions and the numbers of questions of each type on the assessments used by Novick *et al.* (2014) and in the present study. More detailed information about the relation between the questions on the two assessments can be found in Section C of the Supplemental Material. Here, we discuss only the new question types and a change in how the pages were laid out.

New Types of Questions. Novick *et al.*'s (2014) assessment did not test the convergent evolution and evolutionary sequence skills. We added one question for each of these skills to provide a preliminary evaluation of students' success at these aspects of tree thinking. Although one question provides only a weak basis for evaluation, we were constrained by our need not to make the assessment longer. Adding one item for each skill rather than no items was our compromise. Part IV of the phylogenetics laboratory asked students to identify which character state on a cladogram provided evidence for convergent evolution. The assessment asked the converse question: Consider a given character, which appeared twice on the cladogram, and explain why it is shared by two taxa.

For the evolutionary sequence question, students were told to consider the evolutionary relationship between *Sellosaurus* and *Vulcanodon* (see Figure 1) and then were asked "What

[2]A possible concern with this change is that students might have difficulty generating appropriate language for their explanations in a free-response format, whereas they might do considerably better with a multiple-choice format in which they only have to recognize the appropriate language. Examination of the mean proportion correct across the 10 explanation questions on the pretest indicates no support for this hypothesis. For both instructional conditions, these means ranged from 0.02 to 0.16, with an overall mean across the 10 questions of only 0.10.

TABLE 4. The number of test items for each skill that were included on the tree-thinking assessments used in our earlier study and in the present study

Tree-thinking skill	Novick *et al.* (2014)	Present study
Evolutionary relatedness		
Resolved	8	5
Polytomy	4	5
Clades (valid biological groups)		
Evaluate, identify clade	7	5
Nested clades	2	2
Inference	4	8
Convergent evolution	0	1
Evolutionary sequence	0	1
Subsets of the Tree of Life		
Without rotation	10	2
With rotation	0	3
Rotation[a]	0	3
Prior knowledge and tree thinking[b]	2	2
Total number of items	37	37

[a] These questions, which were new for the present study, turned out to be too easy, as students in both instructional conditions did extremely well on them on the pretest (mean = 0.84). Therefore, it does not make sense to include these items in evaluating our tree-thinking instruction.

[b] The prior-knowledge questions were included to take advantage of a captive sample of students to ask questions of interest for another purpose. Thus, we will not discuss them here, and they are not included in the outcome measures reported in this article.

sequence of characters provides evidence for this relationship?" Five alternative character sequences were provided in a multiple-choice format. The incorrect choices mimicked the errors undergraduates made in our earlier research (unpublished data) using a free-response format. This question required students to go beyond what they had been specifically taught, which is another reason for asking only a single question.

Given the problem with Novick *et al.*'s (2014) items intended to assess students' understanding of subsets of the ToL, noted earlier, we wrote five new items for this skill. For two questions, we embedded three particular taxa in three cladograms that included different other taxa. A multiple-choice question asked about the relationships among the three common taxa. For the other three questions, which also included the concept of rotation, students had to indicate whether pairs of cladograms showed the same or different relationships among a subset of the taxa. Scientists often have to reason whether different trees that involve overlapping sets of taxa (e.g., those derived from different data sets) suggest the same patterns of relationships among the common taxa or constitute competing hypotheses about the relationships.

Format of the Test Pages. Given the size of the explanations box, we changed from portrait to landscape mode for the test booklet, with the explanations printed on the right half of the page. All but four of the test questions (two nested clades, two inference) fit on the left side of the page. Unfortunately, most students skipped those four questions, although they answered all the other questions. Accordingly, we did not include those four questions in assessing students' tree thinking. This unfortunately meant we had no data for the marking nested clades skill.

Analysis

Students' responses to the main skill questions were scored as consistent with the evolutionary evidence depicted in the cladogram (i.e., correct; 1) or inconsistent with that evidence (i.e., incorrect; 0). Responses to the follow-up explanation questions were similarly scored as 1 or 0, depending on whether an evolutionarily correct explanation was selected. Proportion correct scores for the individual tree-thinking skills were then computed by averaging the scores for the relevant questions. We also computed a composite tree-thinking score by summing students' scores for the seven skills we were able to measure: evolutionary relatedness: resolved; evolutionary relatedness: polytomy; evaluating/identifying clades; inference; convergent evolution; evolutionary sequence; and subsets of the ToL. This composite equally weights the seven skills and indicates the number of skills on which students were successful. Scores on the composite could range from 0 to 7.

An alternate composite can be computed by averaging (or summing) across all the questions to yield the proportion (or total number) of questions students answered correctly. This composite more heavily weights skills for which we included more items. We believe the composite that equally weights the skills provides a more useful measure of students' competence at tree thinking. Both composites yield the same statistical conclusions.

The individual skill and composite tree-thinking scores were analyzed with separate 2 × 2 mixed analyses of variance (ANOVAs), with instructional condition (business-as-usual vs. experimental; between) and time of test (before [pretest] vs. after [posttest] instruction; within) as factors. In a subsidiary analysis, we used multiple regression to predict composite posttest scores based on year in school, sex, whether the student was a biology major, composite pretest score, and instructional condition. This analysis yielded the same conclusions as does the ANOVA we report in the *Results* section.[3]

RESULTS

The mean pretest and posttest scores for the individual skills and the composite are shown in Table 5. The ANOVA on the composite tree-thinking scores yielded significant main effects of condition and time of test and a significant interaction. Students who received our experimental instruction were more successful at tree thinking than were students who received business-as-usual instruction ($F(1, 115) = 9.03$, $p < 0.01$, MSE = 1.44, $\eta_p^2 = 0.07$), and students did better on the posttest than on the pretest ($F(1, 115) = 58.11$, $p < 0.001$, MSE = 0.68, $\eta_p^2 = 0.34$). The interaction indicated that the improvement from pretest to posttest was significantly larger for students in the experimental instruction condition (1.16) than for those in the business-as-usual condition (0.48; $F(1, 115) = 10.02$, $p < 0.01$,

[3] Year in school, sex, and major reflect the demographic information we have about the students in this study. The multiple regression indicated that only pretest accuracy and instructional condition made significant individual contributions to predicting posttest scores. Thus, the superior posttest performance of students in our experimental instruction condition is maintained when simultaneously controlling for students' prior knowledge of tree thinking and three major demographic variables.

TABLE 5. Mean scores (*SE*s) on the tree-thinking assessment

	Business-as-usual instruction			Experimental instruction		
Tree-thinking skill	Pretest	Posttest	Change	Pretest	Posttest	Change
Composite (Σ seven skills)	1.46	1.94	0.48	1.59	2.76	1.16
	(0.10)	*(0.16)*		*(0.09)*	*(0.17)*	
Evolutionary relatedness: resolved	0.29	0.35	0.06	0.31	0.47	0.16
	(0.03)	*(0.03)*		*(0.03)*	*(0.03)*	
Evolutionary relatedness: polytomy	0.08	0.07	–0.01	0.10	0.19	0.09
	(0.02)	*(0.02)*		*(0.02)*	*(0.03)*	
Evaluating clades	0.18	0.26	0.08	0.27	0.43	0.16
	(0.03)	*(0.03)*		*(0.03)*	*(0.04)*	
Inference	0.22	0.23	0.01	0.23	0.35	0.12
	(0.02)	*(0.02)*		*(0.02)*	*(0.03)*	
Subsets of ToL	0.30	0.32	0.02	0.29	0.39	0.10
	(0.03)	*(0.03)*		*(0.02)*	*(0.03)*	
Convergent evolution	0.13	0.36	0.23	0.10	0.56	0.46
	(0.05)	*(0.07)*		*(0.04)*	*(0.06)*	
Evolutionary sequence	0.25	0.35	0.10	0.31	0.35	0.04
	(0.06)	*(0.06)*		*(0.06)*	*(0.06)*	

MSE = 0.68, $\eta_p^2 = 0.08$). These gain scores yield a Cohen's *d* of 0.59, which represents a medium size effect for the increased effectiveness of our experimental instruction relative to business-as-usual instruction. For the experimental instruction condition alone, Cohen's *d* is 1.10 based on the means and SDs for the pretest and posttest scores and 1.45 taking into account the correlation ($r = 0.37$) between the tests. These *d* values indicate a large effect of our tree-thinking instruction, even though students were unable to study for the posttest.

Most of the individual skill variables show the same pattern found for the composite tree-thinking score (see Table 5). Indeed, the condition-by-time interaction was significant or marginally significant for all of the skills except evaluating clades and evolutionary sequence. For clades, only the two main effects were significant. For evolutionary sequence, there were no significant effects. Except for that skill, the improvement from pretest to posttest was at least twice as large in the experimental condition as in the business-as-usual condition.

Examination of students' responses to the evolutionary sequence question, asked about the cladogram shown in Figure 1, helps to pinpoint students' difficulty in understanding this concept: "Consider the evolutionary relationship between *Sellosaurus* and *Vulcanodon*. What sequence of characters provides evidence for this relationship?" This stem was followed by five multiple-choice options. The correct answer is "elongated neck → reduced number of finger bones → spoon-shaped teeth." The four incorrect alternatives were 1) these characters in the reverse (backward-time) order; 2) the three correctly ordered characters (synapomorphies) plus the unique (i.e., unshared) characters (autapomorphies) specific to each particular taxon: "elongated neck → reduced number of finger bones → spoon-shaped teeth → modified vertebrae (*Sellosaurus*), peg-like teeth (*Vulcanodon*)"; 3) the last synapomorphy and the two autapomorphies; and 4) just the two autapomorphies. On both the pretest and the posttest, nearly half the students (47%) selected the incorrect alternative that included the three synapomorphies plus the two autapomorphies, approximately one-third (32%) chose the correct answer, roughly 10% chose the correct characters in reverse time order and the final synapomorphy plus the two autapomorphies, and ~2% chose the two autapomorphies.

DISCUSSION

Tree thinking, the set of skills used by scientists to understand and reason with the information depicted in diagrammatic representations of the ToL, is an increasingly important aspect of 21st-century science literacy (Thanukos, 2009; Baum and Smith, 2013; Novick and Catley, 2013). As biologists continue to assemble the ToL, with the ultimate goal of understanding the evolutionary relationships among all extant and extinct taxa, the benefits to humankind will continue to accrue in areas as diverse as climate change, health, agriculture, forensics, and biotechnology (e.g., AMNH, 2002; Futuyma, 2004; Thomas *et al.*, 2004; Yates *et al.*, 2004; Davis *et al.*, 2010).

It is of significant concern, therefore, that undergraduate biology students have difficulty engaging in tree thinking even after business-as-usual instruction in phylogenetics in college biology courses, including upper-level courses such as evolution and zoology that have introductory biology as a prerequisite (Sandvik, 2008; Halverson *et al.*, 2011; Catley *et al.*, 2012; Phillips *et al.*, 2012; Dees *et al.*, 2014). Accordingly, to promote this critical aspect of science literacy, we created a curriculum that was guided by research on the nature of students' difficulties engaging in tree thinking. We developed an instructional booklet (Novick *et al.*, 2014), lectures (present study), and a phylogenetics laboratory (present study) to teach undergraduates the nuts and bolts of how to interpret cladograms and engage in tree thinking. We tested the efficacy of these materials, in comparison with that of business-as-usual instruction in phylogenetics, in a large introductory biology course for science majors using a pretest/posttest design.

The Positive Impact of Instruction

Students did very poorly on the pretest (see Table 5), which we expected given that they had had little relevant instruction at that point in the semester. Although the mean composite scores are not different from what would be expected if students guessed randomly on every question, which would yield a composite score of ~1.59, it is unlikely that students responded in this manner. Consider, for example, their pretest responses to the evolutionary sequence question, which presented five multiple-choice alternatives. The pattern of responses, given earlier, is very far from what would be predicted by random guessing ($\chi^2(4, N = 114) = 74.68, p < 0.001$; three students who left the question blank or who gave multiple responses were excluded from the analysis). It is likely that students knew the answers to some pretest questions, were systematically mistaken on other questions, and guessed on the remaining questions.

Comparing students' scores on the pretest and the posttest, it is clear that we succeeded in our primary goal of moving our experimental curriculum from the highly controlled setting used by Novick *et al.* (2014) to regular classroom instruction in introductory biology. We found a medium size effect for the increased effectiveness of our experimental tree-thinking instruction over business-as-usual instruction and a large effect for improvement from pretest to posttest in the experimental instruction condition alone. Moreover, the benefit of our experimental instruction over business-as-usual instruction was observed not just for our composite tree-thinking measure, which summed across the individual skills, but also for most of the individual tree-thinking skills. These results are especially notable, because 1) they are based on performance on an unannounced, closed-book posttest given 7 weeks after the pretest and 5–6 weeks after the main part of the instruction, and 2) students in the business-as-usual condition received a major component of our experimental instruction—namely, the 3-hour phylogenetics laboratory.

Presumably, completing the laboratory, along with the small amount of instruction in phylogenetics provided during their lecture class, is a primary reason why students in the business-as-usual condition showed a significant gain in overall tree-thinking ability from pretest to posttest. However, these gains were restricted to a few specific skills. In particular, these students showed no increase in performance on the evolutionary relationship: polytomy; inference; and subsets of the ToL skills. We defined and illustrated a simple polytomy in the phylogenetics lab but provided no instruction in assessing relative evolutionary relatedness among taxa in such a structure. Similarly, students had to answer a simple inference question in which character data were provided for all but one taxon. We did not discuss the more general case in which inferences are made based on the fact that certain taxa share a more recent common ancestor with each other than do other taxa. The lab did not consider the ToL subsets skill at all. We will say more about the relationship between the polytomy and ToL subsets skills in the section on *The Problem of Reifying Subsets of the Tree of Life*.

A Case of Ineffective Instruction

The one tree-thinking skill for which our experimental instruction did not lead to numerically larger improvement than business-as-usual instruction was evolutionary sequence. As noted earlier, this question required students to go beyond what they had specifically been taught. The frequencies with which students selected the various answer options clarify the primary difficulty they had in understanding this tree-thinking concept. Because students rarely chose the option that had the synapomorphies in reverse time order, it seems they understood the importance and directional nature of the time arrow in evolution (even on the pretest). Indeed, nearly 80% of students selected one of the two options that included all three synapomorphies in the correct order of time. What students failed to understand, even after instruction, is that autapomorphies logically cannot provide evidence to support the relationship of one taxon to another, because those characters must have arisen after the branching event that split their parent taxon. The distinction between synapomorphies (which support relationships) and autapomorphies (which are diagnostic for a specific taxon) needs to be highlighted more clearly than we did in our instruction.

Students' Absolute Level of Performance

The improvement from pretest to posttest was generally at least twice as large in the experimental instruction condition as in the business-as-usual condition. Nevertheless, students' posttest scores were rather low, with means of 1.94 (business-as-usual condition) and 2.76 (experimental instruction condition) out of 7 on the composite tree-thinking measure. This is presumably because the end-of-semester posttest was unannounced and occurred ~5–6 weeks after the primary instruction in phylogenetics.

On the one hand, having an unannounced posttest is a strength of our study, because it demonstrates the effectiveness of our experimental instruction under very stringent test conditions. Students in the experimental condition clearly were able to integrate some of the tree-thinking skills they learned into their permanent skill set. On the other hand, students in college courses rarely receive unannounced tests long after the primary instruction was delivered, and it is well known that students perform much better on tests for which they are able to study in advance (and that additional opportunities for study yield increased learning). Thus, the posttest scores undoubtedly underestimate the effectiveness of our research-based curriculum in the more typical classroom setting in which major tests are announced in advance (e.g., Novick and Catley, in press, 2017).

Evaluating the Phylogenetics Laboratory

A second positive outcome of this study concerns the phylogenetics laboratory we created. Completing this laboratory presumably contributed to the significant increase in the tree-thinking composite score from pretest to posttest for students in the business-as-usual instruction condition.

The laboratory instructors reported that students confidently completed the laboratory, except for some difficulties with part I due to unfamiliarity with some of the character states. Thus, it took them a long time to fill in the character matrix. In a revised version of the laboratory, we filled in some of the more difficult character states for students and provided pictures of the characters to help students know what to look for.

Our research-based laboratory was very well received by the other members of the instructional team for the introductory

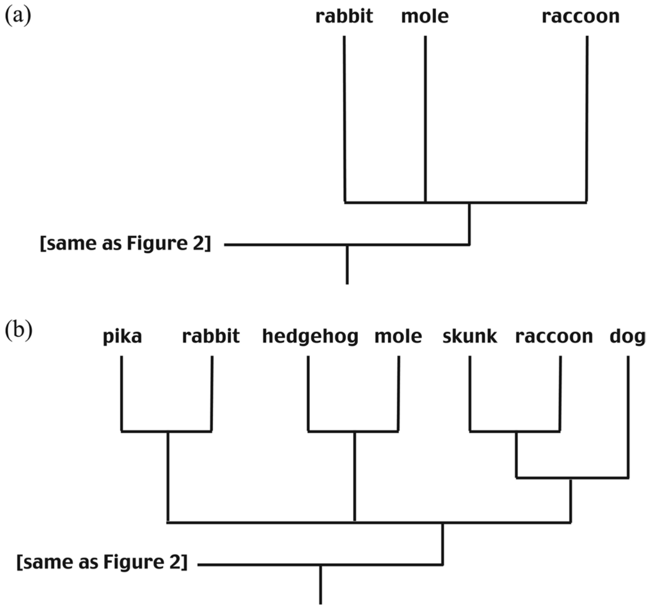

FIGURE 3. **Revisions of the cladogram shown in Figure 2 in which (a) skunk and dog have been removed and (b) pika and hedgehog have been added.**

biology course. As a result, the revised version continues to be the primary means of teaching phylogenetics in a hands-on laboratory environment in the second-semester science majors introductory biology course at K.M.C.'s institution. We have also used this laboratory successfully as part of an extended tree-thinking curriculum in an upper-level organismal biology class (Novick and Catley, in press, 2017).

Evaluating the Revised Tree-Thinking Assessment

A key substantive revision to the assessment instrument was the new ToL subsets questions. This revision was effective, as indicated by the greater improvement from pretest to posttest for the experimental instruction condition ($\Delta = 0.10$) compared with the business-as-usual condition ($\Delta = 0.02$) for these items.

In contrast, our switch from portrait to landscape orientation for the assessments to accommodate the large list of possible explanations provided for each explanation question turned out to be a major problem, because it led students to fail to see (and therefore answer) the four questions that were printed at the top of the right column, immediately above the box of explanations. We have therefore returned to the portrait orientation for the assessments in our subsequent research (Novick and Catley, in press, 2017).

The Problem of Reifying Subsets of the Tree of Life

Looking at the pretest to posttest change scores for students in the experimental instruction condition (see Table 5), it is clear that some skills saw larger improvements than others. Among the skills that were assessed by five to six items (evolutionary relatedness: resolved; evolutionary relatedness: polytomy; identify/evaluate clades; inference; subsets of the ToL), the largest increases ($\Delta = 0.16$) were for evolutionary relatedness: resolved; and identify/evaluate clades. The smallest increases (Δ values of 0.09–0.10) were for evolutionary relatedness: polytomy; and subsets of the ToL. These two skills also had among the lowest posttest scores. Clearly, our instruction was more effective for the former than the latter skills. We hypothesize that students' difficulties with the polytomy and ToL subsets concepts are related. Indeed, we suspect that understanding that any specific cladogram is a subset of the complete, unimaginably large ToL may be the fundamental difficulty students have to overcome in acquiring tree-thinking expertise.

Consider the polytomy relatedness question we asked about the cladogram in Figure 2. Students had to choose whether 1) moles are more closely related to rabbits than to raccoons, 2) moles are more closely related to raccoons than to rabbits, or 3) rabbits, moles, and raccoons are all equally closely related to each other. If skunk and dog were pruned off the cladogram, as shown in Figure 3a, it would be easy to see that rabbits, moles, and raccoons form a polytomy. With those two taxa included, however, students see that raccoons are more closely related to skunks than to rabbits and moles, because raccoons and skunks (and dogs) are joined at a higher level (i.e., share a more recent common ancestor). Rabbits and moles, in contrast, appear at the same level in the cladogram in Figure 2. Of course, we could make rabbits and moles also appear at different levels by adding, for example, pika as the sister group to rabbit and hedgehog as the sister group to mole as shown in Figure 3b. None of these added taxa change the relationship of rabbits, moles, and raccoons to one another because that relationship depends solely on the MRCA of those three taxa, not on the MRCA of any one of those taxa with other taxa. Nevertheless, the visual linkages and hierarchical levels shown in a given cladogram appear to exert a powerful influence on students' reasoning. That is, students seem to reify the visually depicted relationships: Rabbits and moles are at the same level, whereas raccoons join with other taxa one level up, so moles must be more closely related to rabbits than to raccoons. Indeed, all of the students who got this question wrong on the posttest (96% in the business-as-usual condition, 84% in the experimental instruction condition) gave this incorrect answer.

Instead of reifying what they see, students need to keep in mind that there may be an indefinite number of other taxa linked with each depicted taxon. The only constant across changing subsets of taxa is whether two taxa, regardless of where they are located in a particular cladogram, share a more recent common ancestor with each other than with some third taxon. Although this "three-taxon statement" approach was the basic building block of our instructional booklet, our instruction was not completely successful in conveying this key idea. How to improve instruction to address this obstacle to understanding is an important topic for future research.

Limitations of This Study

One limitation of this study stems from its implementation in an extant, large, multisection course with multiple instructors: the business-as-usual and instructional conditions were taught by different professors. Thus, some of the difference in performance between the two conditions logically could be due in part to differences between the instructors. Although we cannot rule out this possibility, we suspect that such differences play a relatively minor role. Students in the two conditions were also taught by different professors in the first half of the semester, before they took the pretest, yet there was no significant difference between their pretest scores. This is consistent with the fact that those two professors presented similar material concerning phylogenetics. After the pretest, the material on phylogenetics presented by the two (new) professors differed both in content and extent. Finally, the phylogenetics laboratory was taught by two different laboratory instructors, and there is no evidence that students in the different laboratory sections performed differently on the posttest.

A second limitation of this study is that the tree-thinking curricula we compared varied in both content and extent, as just noted. Previous research indicates that business-as-usual instruction in phylogenetics, even a whole week of lectures in an upper-level organismal biology or evolution course, is inadequate to promote a high level of competence across the various tree-thinking skills. It is clear, therefore, that a different, more rigorous, and consistent approach is needed. Our study provides initial data on what such a different approach might entail. We supplemented the week of lectures with a short instructional booklet, a phylogenetics laboratory, and the infusion of tree thinking throughout the remainder of the course. It is important to emphasize that the consistent tree-thinking approach promoted during the course was a change in perspective, not a change in content. Students in both conditions used the same textbook and learned about the same taxa. Also, the phylogenetics laboratory we created replaced the laboratory on that topic that was used previously. It represents, we believe, improved content but not additional time on task. In any case, students in both conditions completed that laboratory, and it presumably accounted for part of the increase in tree-thinking scores from pretest to posttest in the business-as-usual condition.

A third limitation stems from the short amount of time between completing data collection for our initial study (Novick *et al.*, 2014) and the beginning of the semester in which K.M.C. was assigned to teach the introductory biology class. For this reason, we were unable to pilot test and validate the new assessment questions and the altered layout of the test pages as extensively as we would have liked.

CONCLUSION

Tree thinking is an indispensable tool for biology majors and a critical component of science literacy generally. Accordingly, we set out to create, implement, and test a research-based tree-thinking curriculum and assessment. Our efforts were very successful in both a highly controlled research setting (Novick *et al.*, 2014) and a large introductory biology course for science majors. In the current study in the introductory biology course, we showed that direct instruction in both a laboratory and lecture setting produced skills that enhanced students' understanding of macroevolutionary patterns and processes after a delay of many weeks.

Further research is needed in several areas of tree thinking: understanding evolutionary relatedness in cladogram topologies that include polytomies, comprehending that any particular cladogram depicts only a subset of the complete ToL, and conceiving of relationships among taxa as being supported by a sequence of shared characters (synapomorphies) reflecting most recent common ancestry. We hypothesize that greater clarity in all of these areas might accrue from a deeper understanding of the concept of subsets as applied to phylogenetic trees.

ACKNOWLEDGMENTS

We thank Emily Schreiber for her help in modifying the instructional booklet and test questions from her honors thesis study for use in the present study. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through grant R305A080621 to Vanderbilt University. The opinions expressed are those of the authors and do not represent views of the institute or the U.S. Department of Education.

REFERENCES

American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education (1999). Standards for Educational and Psychological Testing, Washington, DC: AERA.

American Museum of Natural History (2002). Assembling the Tree of Life: Harnessing Life's History to Benefit Science and Society. Brochure produced for the National Science Foundation based on three NSF Tree of Life workshops held in 1999 and 2000 at Yale University, the University of California, Davis, and the University of Texas, Austin. www.phylo.org/sub_sections/outreach/outreach_b.php (accessed 30 August 2010).

Baum DA, Smith SD (2013). Tree Thinking: An Introduction to Phylogenetic Biology, Greenwood Village, CO: Roberts.

Baum DA, Smith SD, Donovan SS (2005). The tree thinking challenge. Science 310, 979–980.

Bryant HN, Russell AP (1992). The role of phylogenetic analysis in the inference of unpreserved attributes of extinct taxa. Philos Trans R Soc Lond B 337, 405–418.

Campbell NA, Reece JB, Urry LA, Cain ML, Wasserman SA, Minorsky PV, Jackson RB (2008). Biology, 8th ed., San Francisco, CA: Pearson Benjamin Cummings.

Catley KM (1994). Descriptions of new *Hypochilus* species from New Mexico and California with a cladistic analysis of the Hypochilidae (Araneae). American Museum Novitates 3088, 1–27. http://digitallibrary.amnh.org/dspace/handle/2246/4985 (accessed 8 June 2010).

Catley KM (2006). Darwin's missing link: a novel paradigm for evolution education. Sci Educ 90, 767–783.

Catley KM, Novick LR, Funk DJ (2012). The promise and challenges of introducing tree thinking into evolution education. In: Evolution Challenges: Integrating Research and Practice in Teaching and Learning about Evolution, ed. K Rosengren, EM Evans, S Brem, and G Sinatra, New York: Oxford University Press, 93–118.

Catley KM, Phillips BC, Novick LR (2013). Snakes and eels and dogs! Oh, my! Evaluating high school students tree-thinking skills: an entry point to understanding evolution. Res Sci Educ 43, 2327–2348.

Davis CC, Willis CG, Primack RB, Miller-Rushing AJ (2010). The importance of phylogeny to the study of phenological response to global climate change. Philos Trans R Soc Lond B 365, 3201–3213.

Dees J, Momsen JL, Niemi J, Montplaisir L (2014). Student interpretations of phylogenetic trees in an introductory biology course. CBE Life Sci Educ 13, 666–676.

Eddy SL, Crowe AJ, Wenderoth MP, Freeman S (2013). How should we teach tree-thinking? An experimental test of two hypotheses. Evol Educ Outreach 6, 13.

Futuyma DJ (2004). The fruit of the tree of life. In: Assembling the Tree of Life, ed. J Cracraft and MJ Donoghue, New York: Oxford University Press, 25–39.

Gregory TR (2008). Understanding evolutionary trees. Evol Educ Outreach 1, 121–137.

Halverson KL, Pires CJ, Abell SK (2011). Exploring the complexity of tree thinking expertise in an undergraduate systematics course. Sci Educ 95, 794–823.

Hennig W (1966). Phylogenetic Systematics, Urbana: University of Illinois Press.

Kaplan RM, Saccuzzo DP (2009). Psychological Testing: Principles, Applications, and Issues, 7th ed., Belmont, CA: Wadsworth.

McLaurin DC, Halverson KL, Boyce CJ (2013). Using manipulative models to develop tree-thinking. Biol Int 54, 108–121.

Meir E, Perry J, Herron JC, Kingsolver J (2007, September). College students' misconceptions about evolutionary trees. Am Biol Teach Online 69, e71.

Meisel RP (2010). Teaching tree-thinking to undergraduate biology students. Evol Educ Outreach 3, 621–628.

National Research Council (2012). Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering, Washington, DC: National Academies Press.

Novick LR, Catley KM (2013). Reasoning about evolution's grand patterns: college students' understanding of the tree of life. Am Educ Res J 50, 138–177.

Novick LR, Catley KM (2017). Evaluating the effectiveness of a multi-faceted tree-thinking curriculum in an upper-level organismal biology course. J Biol Educ (*in press*).

Novick LR, Schreiber EG, Catley KM (2014). Deconstructing evolution education: the relationship between micro- and macroevolution. J Res Sci Teach 51, 759–788.

Phillips BC, Novick LR, Catley KM, Funk DJ (2012). Teaching tree thinking to college students: it's not as easy as you think. Evol Educ Outreach 5, 595–602.

Sandvik H (2008). Tree thinking cannot taken for granted: challenges for teaching phylogenetics. Theory Biosci 127, 45–51.

Singer F, Hagen JB, Sheehy RR (2001). The comparative method, hypothesis testing & phylogenetic analysis—an introductory laboratory. Am Biol Teach 63, 518–523.

Smith JJ, Cheruvelil KS (2009). Using inquiry and tree-thinking to "march through the animal phyla": teaching introductory comparative biology in an evolutionary context. Evol Educ Outreach 2, 429–444.

Smith JJ, Cheruvelil KS, Auvenshine S (2013). Assessment of student learning associated with tree thinking in an undergraduate introductory organismal biology course. CBE Life Sci Educ 12, 542–552.

Thanukos A (2009). A name by any other tree. Evol Educ Outreach 2, 303–309.

Thomas CD, Cameron A, Green RE, Bakkenes M, Beaumont LJ, Collingham YC, Erasmus BFN, Ferreira de Siqueira M, Grainger A, Hannah L, *et al.* (2004). Extinction risk from climate change. Nature 427, 145–148.

Witmer LM (1995). The extant phylogenetic bracket and the importance of reconstructing soft tissues in fossils. In: Functional Morphology in Vertebrate Paleontology, ed. J Thomason, New York: Cambridge University Press, 19–33.

Yates TL, Salazar-Bravo J, Dragoo JW (2004). The importance of the tree of life to society. In: Assembling the Tree of Life, ed. J. Cracraft and MJ Donoghue, New York: Oxford University Press, 7–17.