

Article

A Deliberate Practice Approach to Teaching Phylogenetic Analysis

F. Collin Hobbs,^{*†} Daniel J. Johnson,^{*‡} and Katherine D. Kearns[§]

^{*}Department of Biology and [§]Center for Innovative Teaching and Learning, Indiana University, Bloomington, IN 47405

Submitted March 3, 2013; Revised September 21, 2013; Accepted September 24, 2013
Monitoring Editor: Kenneth Raymond Miller

One goal of postsecondary education is to assist students in developing expert-level understanding. Previous attempts to encourage expert-level understanding of phylogenetic analysis in college science classrooms have largely focused on isolated, or “one-shot,” in-class activities. Using a deliberate practice instructional approach, we designed a set of five assignments for a 300-level plant systematics course that incrementally introduces the concepts and skills used in phylogenetic analysis. In our assignments, students learned the process of constructing phylogenetic trees through a series of increasingly difficult tasks; thus, skill development served as a framework for building content knowledge. We present results from 5 yr of final exam scores, pre- and postconcept assessments, and student surveys to assess the impact of our new pedagogical materials on student performance related to constructing and interpreting phylogenetic trees. Students improved in their ability to interpret relationships within trees and improved in several aspects related to between-tree comparisons and tree construction skills. Student feedback indicated that most students believed our approach prepared them to engage in tree construction and gave them confidence in their abilities. Overall, our data confirm that instructional approaches implementing deliberate practice address student misconceptions, improve student experiences, and foster deeper understanding of difficult scientific concepts.

INTRODUCTION

One purpose of higher education is to engage students in critical thinking: to move students beyond memorization of terms, so they can construct relationships between concepts, apply appropriate frameworks to problem solving, and critique conclusions (Hackling and Garnett, 1992; National

Research Council [NRC], 2000; Knight and Wood, 2005; Tanner and Allen, 2005). Students who learn to think like experts can see patterns in information, infer meaning in those patterns, and transfer that meaning to novel situations and problems (NRC, 2000; Wiggins and McTighe, 2005). These expert skills are constructed at multiple scales in a student’s education. At the curricular level, a learning progression spanning multiple years can help teams of instructors uncover novice explanations of phenomena and intentionally advance students’ expert-level, scientific explanations (NRC, 2007; Duncan and Hmelo-Silver, 2009; Mohan *et al.*, 2009). At the course level, instructors must implement activities and assignments to address the naïve theories, preconceptions, and misconceptions that can impede understanding (Wiggins and McTighe, 2005). In addition, students must pay focused attention to and practice extensively with several layers of knowledge: factual (terminology), conceptual (categories), procedural (skills, methods, and criteria), and metacognitive (self-knowledge; Ericsson *et al.*, 1993; Anderson *et al.*, 2001).

Expert-level thinking is necessarily discipline specific (Middendorf and Pace, 2004; Bauer-Dantoin, 2009; Chick

DOI: 10.1187/cbe-13-03-0046

Present addresses: [†]Department of Biology, Huntington University, Huntington, IN 46750; [‡]Department of Evolution, Ecology, and Organismal Biology, Ohio State University, Columbus, OH 43210.

Address correspondence to: F. Collin Hobbs (chobbs@huntington.edu).

© 2013 F. C. Hobbs *et al.* CBE—Life Sciences Education © 2013 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution-Noncommercial-Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

“ASCB[®]” and “The American Society for Cell Biology[®]” are registered trademarks of The American Society for Cell Biology.

et al., 2009). In the field of evolution, experts represent and interpret evolutionary relationships through phylogenetic trees, diagrams that structure our knowledge of biological diversity and represent evolutionary relationships among taxa (“tree-thinking”; see Baum *et al.*, 2005; Baum and Offner, 2008; Meir *et al.*, 2007). Yet undergraduate students, even those with advanced course work in biology, struggle to construct, interpret, and make comparisons among phylogenetic trees (Anderson *et al.*, 2002; Baum *et al.*, 2005; Shtulman, 2006; Meir *et al.*, 2007; Perry *et al.*, 2008; Lents *et al.*, 2010; Halverson *et al.*, 2011). Such misunderstandings and lack of tree-thinking skills can diminish students’ perception of science and the scientific process in general; furthermore, these difficulties can impede the ability of American students to contribute meaningfully to the ongoing discussion surrounding the teaching of evolution in the U.S. education system (Baum *et al.*, 2005; Lents *et al.*, 2010; Berkman and Plutzer, 2011).

Instructional interventions to improve students’ tree-thinking skills have incorporated a variety of course-level teaching strategies, research designs, student populations, and learning assessments (Goldsmith, 2003; Burks and Boles, 2007; Perry *et al.*, 2008; Lents *et al.*, 2010). These previous studies provide support for targeted exercises that tackle specific evolution misconceptions and thereby improve students’ knowledge about phylogenetics and basic tree-reading skills. However, research from the cognitive sciences suggests that such instruction does not replace the naïve theories held by students; instead, the two tend to coexist (Shtulman and Valcarcel, 2012). This observation may result, in part, from the emphasis on “stand-alone” or “one-shot” exercises in current pedagogy practice.

Deliberate practice may better cultivate expert-level understanding of a concept (Ericsson *et al.*, 1993; van Gog *et al.*, 2005; Cepeda *et al.*, 2009). The process of deliberate practice involves allocating multiple sessions for relevant instruction, increasingly complex activities, and frequent revision and feedback opportunities as a means to improve students’ performance (Ericsson *et al.*, 1993; van Gog *et al.*, 2005). While increasing time on task alone has little effect on student academic achievement (Plant *et al.*, 2005), students who engage in deliberate practice have higher performance in problem-oriented courses, such as accounting and chemistry (Crippen and Brooks, 2009; Yu, 2011). Similar results have been observed in college introductory biology courses. For example, Lord (1997) improved student learning in an introductory biology course using constructivist approaches that built incrementally upon students’ prior knowledge. Similarly, both Freeman *et al.* (2011) and Haak *et al.* (2011) improved students’ content knowledge in introductory biology class settings using highly structured approaches that emphasized consistent student preparation before class sessions; repeated content exposure; frequent low-risk assessments; and regular, required in-class exercises.

We implemented a deliberate practice instructional approach to improve our students’ tree-thinking skills in an upper-level plant biology course. Specifically, we redesigned a set of pre-existing laboratory assignments to include five new, increasingly complex, tree-thinking exercises that focused on developing the skill of tree construction. These new exercises placed content acquisition—the learning of relevant vocabulary and concepts—within a framework of skill development, while requiring little additional class time. While our

new teaching materials may be useful in college-level biology courses that discuss evolutionary trees, the general approach used in deliberate practice is relevant to a much broader context in higher education. Specifically, we were interested in the impact of our application of deliberate practice on the ability of our students to interpret relationships within phylogenetic trees, compare among trees, and construct trees. In this paper, we document the multiple measures of success of our exercises in improving expert-level tree-thinking abilities.

MATERIAL AND METHODS

Study Context

Our investigation was conducted in the upper-level, lecture-lab biology course B300, Vascular Plants. This course is offered each Spring semester by the biology department at a large, midwestern public research university. Course enrollment varies from 60 to 70 students each semester. The course is open to both majors and nonmajors, although typically 90% or more of the students are biology majors. In this plant systematics course, students study the morphologies, life cycles, classification, and economic importance of the major vascular plant taxa. All students attend the course lectures taught by the faculty instructor and then participate in one of the four lab sections cotaught by pairs of graduate teaching assistants (GTAs). These weekly, 3-h lab sections are intended to engage students with hands-on experience with the organisms, concepts, and skill sets discussed in the lecture.

The content of the course is organized around an evolutionary framework, both in the progression in which taxa are presented and in how knowledge of the various traits and features is structured. In lecture, students see phylogenetic trees on a daily basis and are required to memorize the inferred evolutionary relationships of major taxa. The laboratory sections focus on developing both content knowledge, such as vocabulary and concepts, and skill-based knowledge, such as constructing and interpreting phylogenetic trees. Historically, content knowledge related to phylogenetic analysis was introduced through a set of three homework assignments consisting of questions based on readings from the course textbook (Simpson, 2010), followed by in-lab discussions of each assignment led by the GTAs (Figure 1A). Instructional time for developing skill-based knowledge was restricted to a single lab (lab 10, Figure 1A). In this lab, students were expected to learn how to construct evolutionary trees both by hand and by using the cladistic software package PAUP* (Sinauer Associates, Sunderland, MA), interpret within-tree relationships, and compare and contrast between competing trees.

We observed several problems that resulted from separating the content and skill-based components of phylogenetic analysis. This separation of content and skills was unlike the instructional approach taken with most other lab topics, which emphasized the concurrent development of both forms of knowledge. For example, while attempting to identify a plant specimen to the species level, students also learned the names and functions of new plant structures relevant to the identification process. Because the laboratory homework assignments focused entirely on terminology and basic concepts, the in-lab discussions were largely dry and unengaging. Further, the process of constructing realistic evolutionary

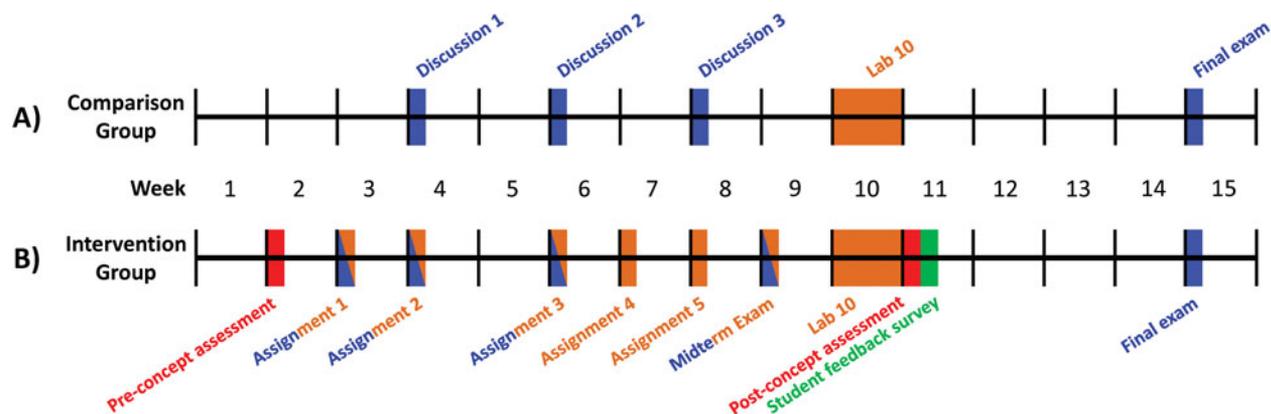


Figure 1. Weekly schedule of B300 Vascular Plants lab sessions before (A) and after (B) the implementation of our new assignments. Items coded blue represent content-based information, while orange represents skill-based information. The timing of the pre- and postconcept assessments (red), student feedback survey (green), midterm exam, and final exam are also shown.

trees is complex, and students struggled to acquire this skill within the confines of a single 3-h laboratory session. As a result, it was not uncommon for students, as well as GTAs, to comment that lab 10 was “the worst lab of the semester.” We also noted that students performed poorly on final laboratory exam questions related to phylogenetic analysis. Finally, students were never formally assessed for their ability to construct phylogenetic trees outside of the lab 10 exercises. We identified this problematic instructional approach as an opportunity to assess the effectiveness of deliberate practice approaches to teaching tree-thinking concepts and skills, while simultaneously improving the course.

Beginning in 2010, we implemented five redesigned laboratory assignments in Vascular Plants (Figure 1 and Table

1). The new assignments were used across all laboratory sections to maintain uniformity in the course and to not put any sections at a disadvantage. Thus, all students enrolled in the course for the 2010 and 2011 Spring semesters form our intervention group (eight sections, 134 students). Our comparison group comprises those students enrolled in the course during the Spring semesters of 2007–2009, prior to the implementation of the new exercises (nine sections, 124 students). During the data-collection period, the same faculty instructor taught the lecture portion of this course; however, eight different GTAs taught the lab sections, including two of the authors (F.C.H. and D.J.J.). Our study was considered exempt by our institution’s human research protections program, as we used anonymous student course work in our analyses, which were

Table 1. Our approach to teaching phylogenetic analysis in the comparison and intervention groups

Category	Comparison group	Intervention group
Theoretical framework		
Teaching approach ^a	Content-centered	Content learning through deliberate practice of skills
Learning objectives ^b	Remember	Application and analysis
Student preparation per assignment ^c		
Textbook readings (×3)	~10 pages	~10 pages
Discussion questions (×3)	~5 questions	~5 questions
Tree-thinking exercises (×5)	N/A	1 exercise
Time management in lab ^c		
Discussion questions (×3)	10 min	10 min
Tree-thinking exercises (×5)	N/A	10 min
Other lab activities	2 h, 40 min	2 h, 30 min
Total lab time	2 h, 50 min	2 h, 50 min
Assessments		
Midterm exam	No phylogenetic questions	Three-part question, 20% of exam grade
Final exam	7 questions, 14% of exam grade	7 questions, 14% of exam grade
Pre/postassessments	N/A	Yes, not graded
Student feedback survey	N/A	Yes, not graded

^aEricsson *et al.*, 1993; Knight and Wood, 2005; Tanner and Allen, 2005.

^bAnderson *et al.*, 2001.

^c“Student preparation . . .” and “Time management . . .” describe the activities students performed for each lab session in which we addressed a phylogenetic analysis assignment (discussions 1–3 and assignments 1–5 in Figure 1). Numbers in parentheses indicate the number of times an activity was repeated in the course.

conducted after the end of the semester (IRB study #08-13645). In compliance with our institutional requirements for an exempt investigation, students in the intervention group were informed at the beginning of the semester about the study through information sheets, which were also posted on the course website. These information sheets did not contain specific information about what components of the course were under examination.

Description of the Teaching Intervention

Our set of five new assignments progressed through the steps of phylogenetic tree construction, beginning with basic skills and culminating with relatively complex tasks (see Supplemental Material). The first three assignments each contained two components: 1) a set of revised content-focused discussion questions; and 2) a newly implemented skill-based exercise. The final two assignments consisted entirely of new skill-based exercises. The exercises in all five assignments were intentionally designed to use the taxa and character states encountered by our students in lecture and lab. Thus, our exercises had the added benefit of familiarizing our students with content they would need for nonphylogenetic questions on the lecture and lab exams.

The discussion question component of the assignments was based on assigned readings from the course textbook (Simpson, 2010) and covered the terminology and concepts related to the skills of constructing and interpreting phylogenetic trees. The sets of discussion questions used in the first three assignments existed prior to our teaching intervention and were used by all students in the comparison group; however, several of the discussion questions were revised prior to implementing the assignments in our intervention group (Figure 1A). The wording and content of these questions were modified to emphasize higher-order cognitive skills (Anderson *et al.*, 2001; Crowe *et al.*, 2008), which we expected would better engage our students and enhance the in-lab discussions (Table 1).

The skill-based component of the assignments was developed for our intervention group and formed the core of our deliberate practice approach to teaching phylogenetic analysis. This part of the assignment presented a hands-on task for the student to complete, followed by a set of reflective questions. These tasks became increasingly complex and comprehensive as the assignments progressed. Students first learned how to build a character-by-taxon matrix, which was then used to organize the taxa and character traits needed to construct a phylogenetic tree. In the second exercise, students completed a larger character-by-taxon matrix, and then mapped the character traits in this matrix onto a blank “skeleton” tree that was provided to them. This exercise reinforced the relationship between character traits and the evolutionary patterns described by phylogenetic trees. In the third and fourth exercises, students learned two different techniques to construct phylogenetic trees by hand (“base-to-tips” and “tips-to-base”), using the information contained in character-by-taxon matrices and extant outgroups as reference taxa. These two exercises also introduced homoplasy (the independent origin of a trait in two or more lineages), polytomies (nodes that have three or more branches), and the possibility that there may be more than one most-parsimonious solution for a given data set. In the final exercise, students constructed a relatively large and complex tree

from a matrix of 25 hypothetical DNA sequence positions for 13 taxa. This exercise required the students to transfer the logic they used in constructing trees from morphological traits to a new situation, and challenged them to apply their skill-based knowledge to organize a large quantity of data. In these exercises, students worked with phylogenetic trees that contained only information about an inferred evolutionary pattern of descent. In other words, the branch lengths of these trees did not convey information about the amount of time or the number of character state changes that had occurred, thereby simplifying the process of constructing trees by hand and allowing students to focus on the evolutionary relationships, or topology, described in the tree.

While completing each assignment, students had multiple opportunities to further their understanding of the concepts and skills through individual work, small-group and class-wide discussion, and revision. Students received each assignment 1 wk in advance to allow them to work through it in their own time. We encouraged student preparation for discussion by requiring them to bring a printed hard copy of their answers to the lab. During lab, students met in small groups for 10–15 min to discuss their answers to the discussion questions and the exercise, after which the GTAs led a class-wide discussion on the assignment that emphasized student presentations. We allowed students to handwrite modifications to their answers during the discussion time. The assignments were submitted for grading following the discussion. The assignments were low-risk, comprising approximately 4% of a student’s total course grade. Our new assignments required little additional in-class time to be devoted to phylogenetic analysis, as time had been previously set aside for discussion questions in the course and students completed the new assignments as homework (Table 1).

Assessments

Performance on Exam Questions. We compared student performance on the final laboratory exam between the intervention group (2010–2011, $n = 134$) and the comparison group (2007–2009, $n = 124$) to assess the impact of our new assignments on overall student performance. This assessment occurred 5 wk after the phylogenetic activities in lab 10. The final laboratory exam for B300 was designed many years prior to our study by the faculty instructor and included seven questions related to phylogenetic analysis that remained unchanged between 2007 and 2011 (see Supplemental Material). Three of the phylogenetic questions on the final exam primarily dealt with understanding and interpreting relationships within trees. The other four phylogenetic questions on the final exam focused primarily on the ability of students to compare trees. Of the nonphylogenetic exam questions, the 25 that had not changed over the study period were used in our analysis. All questions on the exam followed either a multiple-choice or short-answer format. We classified the phylogenetic content questions into the cognitive process categories of Bloom’s revised taxonomy (Anderson *et al.*, 2001; Krathwohl, 2002). One “remember” question required only the ability to remember the definition of a term. Four “apply” questions used students’ understanding of a term to apply it to a given problem. Finally, two “analysis” questions focused on the ability of students to organize and differentiate between the provided materials. Our classification is consistent with that used by Crowe *et al.* (2008).

Performance for each final exam question was defined by the number of correct or incorrect answers in either the intervention (2010–2011) or comparison (2007–2009) group. We used logistic regression to statistically model student performance on these exam questions, with comparison or intervention group as the predictor variable for both the overall performance and question-by-question performance. Logistic regression provides estimates similar to linear regression with a constant term and coefficients for each predictor variable. The coefficients can be transformed into odds ratios by exponentiation of the coefficient. The odds ratio for a binary predictor indicates how much more likely a successful outcome is under one condition compared with another condition. For our purposes, an odds ratio significantly greater than one for the intervention group was considered to be an indication of a significant effect of our teaching intervention. We also performed a logistic regression analysis with the nonphylogenetic questions to test for significant differences in student achievement between the comparison and intervention group independent of our teaching intervention.

To further gauge student learning of tree construction skills in the intervention group, we created a new item for the second laboratory exam, which occurred after the completion of our five assignments (Figure 1B). This new item is a phylogenetic tree construction problem with several questions covering terminology, within-tree metrics, and construction of a phylogenetic tree from a provided character-by-taxon matrix. This question was challenging, with the matrix consisting of seven taxa and an extant outgroup and the most parsimonious answer containing two instances of homoplasy, one of which could be constructed using either convergence or reversal of character traits. Student responses were assessed according to tree topology and correct placement of character state changes.

Pre- and Postconcept Assessment. We designed a pre- and postconcept assessment that was implemented with the intervention group to assess how our students' abilities to interpret and construct evolutionary trees developed over the course of a semester (see Supplemental Material). The design of our concept assessment was informed by both Anderson *et al.*'s (2002) multiple-choice concept inventory and Meir *et al.*'s (2007) open-ended assessment instrument. Our concept assessment contained five questions that addressed key misconceptions commonly held by undergraduate students when interpreting evolutionary trees and two questions that assessed their ability to perform the skill-based tasks of deducing ancestral traits and constructing a basic phylogenetic tree (Baum *et al.*, 2005; Meir *et al.*, 2007). Students in the intervention group completed the preconcept assessment during the second week of classes, before the first of the phylogenetic assignments. The postconcept assessment was conducted in week 11 of the course, following the completion of the five assignments and the phylogenetic analysis lab (Figure 1B). The assessment did not contribute to students' grades in the course, although participation was encouraged by setting aside class time to take the assessment. Only those students who took both the pre- and postconcept assessments were included in our analysis ($n = 121$). Two versions of the assessment were created in order to minimize

test effects of student familiarity (Kennedy, 2005). We designed the versions such that each question in version A had a homologue in version B that addressed the same topic and was nearly identical in content, but differed in wording and graphics. The order of questions in the two versions was also rearranged. The two versions were administered in a cross-over design such that two class sections used version A for the preconcept assessment and version B for the postconcept assessment, while the other two sections did the reverse. A post hoc test for difference in mean score between versions A and B of the preconcept assessment indicates that there was no significant difference in mean student score ($t = -0.666$, p value = 0.2537); therefore, we considered versions A and B identical in subsequent analyses.

A scoring rubric (see Supplemental Material) was developed by the authors (F.C.H. and D.J.J.) after examining 10% of the 2010 preconcept assessments. The rubric defined categories of possible answers for each question and associated scores; answers that demonstrated greater comprehension of the concept or skill received higher scores. We intensively analyzed the tree construction question (Q.7) by scoring responses in each of the four diagnostic categories used by Meir *et al.* (2007): tree topology, placement of common ancestor, correct placement of character traits, and alignment of extant taxa at branch tips. Thus, our concept assessment consisted of 10 pre/post items, five focusing on conceptual knowledge and five on skill-based knowledge. All student responses to the pre- and postconcept assessments from 2010 and 2011 were scored separately by the authors (F.C.H. and D.J.J.) using the rubric. Interrater reliability was determined by calculating Cohen's kappa (Cohen, 1960) for each year's pre- and postconcept assessments. The 2010 preassessment kappa value was 0.80, with 92.1% agreement between raters. The 2010 postassessment kappa was 0.89, with 93.4% agreement between raters. The following year, the preconcept assessment kappa was 0.83 (92.3% agreement), and the postconcept assessment kappa was 0.78 (85.2% agreement). Any student responses that were scored differently were reviewed by both authors in order to arrive at a consensus. Using a single-case research design (Kennedy, 2005), each student's pre- and postconcept assessment responses were matched, and the scores for each question were compared using the normalized change metric (c) (Marx and Cummings, 2007). While similar to normalized gain (Hake, 1998), the normalized change metric better quantifies the relative gains or losses made by each student by excluding the scores of individual students who received full or no credit on a given question in both the pre- and postconcept assessments (Marx and Cummings, 2007).

Student Feedback Surveys Student feedback surveys were conducted in 2010 and 2011 to assess students' perceptions regarding their preparation for the phylogenetic analysis activities in lab 10. The surveys were conducted anonymously in class the week after lab 10 and did not contribute in any way to student grades. The survey consisted of two questions, neither of which referred directly to the phylogenetic exercises we implemented:

- What prepared you the most for the phylogenetic exercises of lab 10?

Table 2. Summary of logistic regression analysis of student answers on final lab exam questions^a

Question	Model terms	Coeff.	95% CI	SE	Z value	p value	Odds ratio
All phylogenetic	Constant	1.09	(0.94, 1.25)	0.08	13.96	< 0.001*	1.56
	Intervention	0.44	(0.22, 0.67)	0.12	3.82	< 0.001*	
All nonphylogenetic	Constant	1.61	(1.52, 1.70)	0.05	33.39	< 0.001*	0.96
	Intervention	-0.04	(-0.17, 0.09)	0.07	-0.65	0.517	
2a	Constant	-0.26	(-0.62, 0.09)	0.18	-1.43	0.152	2.32
	Intervention	0.84	(0.35, 1.35)	0.26	3.30	< 0.001*	
2b	Constant	0.71	(0.34, 1.09)	0.19	3.70	< 0.001*	1.16
	Intervention	0.15	(-0.38, 0.68)	0.27	0.56	0.579	
11a	Constant	1.01	(0.63, 1.43)	0.20	4.99	< 0.001*	0.85
	Intervention	-0.16	(-0.71, 0.38)	0.28	-0.58	0.564	
17a	Constant	1.65	(1.19, 2.16)	0.24	6.75	< 0.001*	3.49
	Intervention	1.25	(0.39, 2.22)	0.46	2.73	0.006*	
17b	Constant	3.70	(2.72, 5.10)	0.58	6.33	< 0.001*	3.30
	Intervention	1.19	(-0.88, 4.21)	1.16	1.03	0.304	
20a	Constant	1.43	(1.00, 1.9)	0.23	6.28	< 0.001*	1.55
	Intervention	0.44	(-0.23, 1.11)	0.34	1.28	0.200	
20b	Constant	1.28	(0.87, 1.72)	0.22	5.88	< 0.001*	2.83
	Intervention	1.04	(0.33, 1.80)	0.37	2.79	0.005*	

^aCoeff. = estimate of the model term or coefficient; 95% CI = 95% confidence interval; SE = and SE of the coefficient; Z value = the test statistic. p values < 0.05 are marked with an asterisk. Odds ratio = the exponentiation of the coefficient by the base *e* and indicates how much more likely a successful outcome is in the intervention group.

- In what ways did you feel unprepared for the phylogenetic exercises of lab 10?

Of the 134 students in the intervention group, 107 students provided responses to the feedback survey. Student responses to the survey were grouped into categories that were created a posteriori by F.C.H. and reviewed by D.J.J. In instances in which a student provided multiple responses to a question, each response was grouped separately into a category. For the first question, 127 responses were categorized; for the second question, 103 responses were categorized. The frequency of responses in each category was calculated.

RESULTS

Performance on Exam Questions

Students in the intervention group, those who experienced the new laboratory assignments, performed significantly better on final laboratory exam questions related to phylogenetics than did students in the comparison group (Table 2 and Figure 2). Results of the logistic regression indicate that students in the intervention group were 1.56 times more likely to answer questions correctly about phylogenetic content than students in the comparison group. For nonphylogenetic questions, the log odds between comparison and intervention

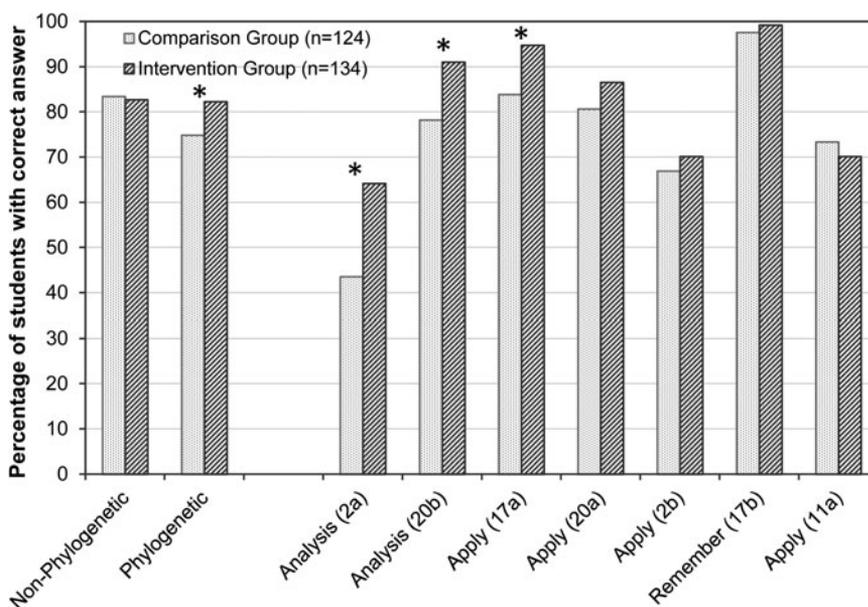


Figure 2. Final exam results for the comparison (2007–2009) and intervention (2010–2011) groups. The first two pairs of bars present the average performance for all exam questions grouped by phylogenetic or nonphylogenetic content. The following seven pairs of bars are individual phylogenetic exam questions. Phylogenetic questions are classified according to Bloom's revised taxonomy (Anderson *et al.*, 2001). Asterisks indicate statistically significant ($\alpha = 0.05$ level) differences between the comparison and intervention groups.

Table 3. Results of the normalized change analysis of the pre- and postconcept assessment, 2010 and 2011 combined

Pre- and postconcept assessment questions	C_{ave}	SEM_c	n_c	Pre-int. (%)	Pre/post 100 (%)	Pre/post 0 (%)	Cohen's d
Concepts							
1. Magnitude of evolutionary change	0.113*	0.044	119	42.7	0.8	0.8	0.413
2. Direction of time on a tree	0.048	0.223	21	82.6	74.4	8.3	0.022
3. Impact of homoplasy on shared traits, relatedness	0.433*	0.085	67	81.5	44.6	0	0.296
4. Ability of trees to rotate at nodes	0.273	0.304	11	94.2	90.9	0	0.120
5. Relatedness depends on common ancestry	0.032	0.041	92	67.4	23.1	0.8	-0.080
Skills							
6. Deducing ancestral traits	0.227*	0.046	103	49.0	14.9	0	0.329
7. Constructing a basic phylogenetic tree							
a. Constructing the branching pattern (topology)	0.020	0.141	51	72.7	52.1	5.8	0.003
b. Placement of the extinct common ancestor	-0.714*	0.084	70	81.0	31.4	10.7	-0.171
c. Mapping character state changes	0.470*	0.062	80	39.7	17.4	16.5	0.141
d. Aligning extant taxa at branch tips	0.911*	0.062	45	55.4	53.7	9.1	0.117

C_{ave} = average normalized change, which can range from -1 to 1; results significantly different from zero ($p < 0.05$) are marked with asterisks. SEM_c = SEM for C_{ave} ; n_c is the number of student scores (out of 121 possible) used in the C_{ave} calculation. "Pre-int." is the overall average score in the preintervention concept assessment. "Pre/post 100" and "Pre/post 0" show the number of students who provided perfect or completely incorrect answers, respectively, on both the pre- and postassessments. Cohen's d = the raw effect size for each question where larger numbers indicate less overlap in the distribution of pre/post data.

groups were not significantly different (Table 2), indicating that our new assignments were the likely factor explaining the higher performance of the intervention group.

With respect to individual phylogenetics questions on the final laboratory exam, the difference in correct answers between the intervention and comparison groups varied considerably (Figure 2). While our intervention group had a greater proportion of the students with correct answers than the comparison group on all but one question, we saw significantly different log odds only between comparison and intervention groups in three of the seven phylogenetics questions (Table 2). These three questions required analysis or application of knowledge concerning the interpretation of evolutionary relationships on phylogenetic trees. For example, the most substantial improvement in odds of a correct answer involved comparing and contrasting the traits of taxa across six similar phylogenetic trees.

Most students in the intervention group were well prepared to construct a phylogenetic tree in a high-stakes situation. Of the 135 intervention group students who took the second laboratory exam, 57.8% of students received full credit for constructing an error-free phylogenetic tree. Almost all students (93.3%) received 75% or more of the possible points, indicating that intervention group students were capable of constructing complex phylogenetic trees.

Pre- and Postconcept Assessment

Results from the pre- and postconcept assessment indicate that intervention group students improved during the semester in their conceptual understanding of phylogenetic trees and in their skill at constructing and interpreting phylogenetic trees. The overall C_{ave} score for all questions was 0.139 ± 0.006 , indicating an overall improvement between pre- and postconcept assessments. Of the five assessment items that focused on conceptual knowledge, two items showed significant gains in understanding over the course of the semester. These items addressed the common misconceptions that a

straight line in a phylogenetic tree is equal to no change in the taxa (Q.1), and that proximity of branch tips implies close relationships between taxa (Q.3) (Table 3 and Figure 3). A majority of students maintained perfect scores between the pre- and postconcept assessment questions concerning the direction of time on a phylogenetic tree (Q.2) and the ability of trees to rotate at their nodes (Q.4; 74.4% and 90.9%, respectively), resulting in small effective sample sizes ($n_c = 21$ and 11, respectively); the students whose scores changed between pre- and postconcept assessments on these questions showed improvement on average, but not significantly. Finally, despite a large effective sample size ($n_c = 92$), students did not significantly improve on the concept question regarding the dependence of relatedness on common ancestry question (Q.5).

Students showed significant improvements in three skill-based items: inferring ancestral traits when provided with information about the traits of extant taxa (Q.6), mapping

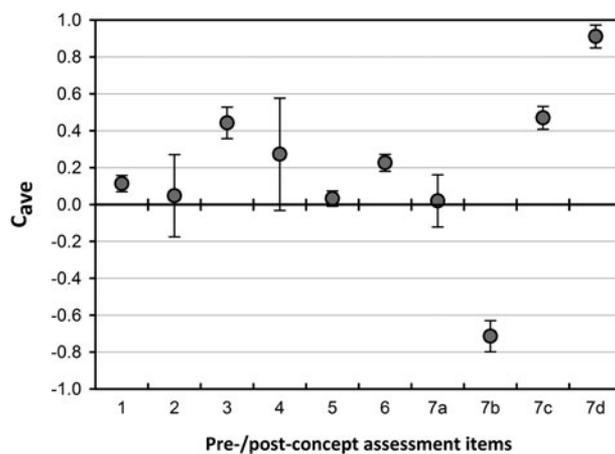


Figure 3. Normalized change scores from the pre- and postconcept assessment. Error bars indicate the SEM for each question.

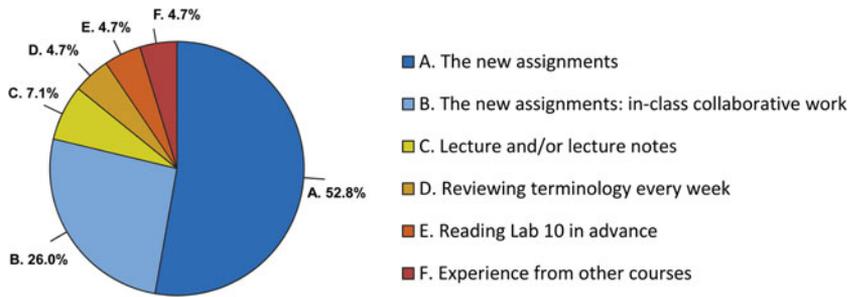


Figure 4. Summary of student responses to the question: “What prepared you the most for the phylogenetic exercises of lab 10?” ($n = 127$; combined 2010 and 2011).

character state changes while constructing trees (Q.7c), and aligning extant taxa at branch tips while constructing trees (Q.7d; Table 3 and Figure 3). Many students (52.1%) correctly constructed the branching pattern of the most parsimonious tree on both the pre- and postconcept assessments (Q.7a); however, those students whose scores did change did not show significant improvement. Finally, although a substantial percentage of students (31.4%) continued to place the extinct common ancestor correctly at the base of the tree, students whose scores did change on this item showed a significant decrease in ability in the postconcept assessment.

Student Feedback Surveys

Students in the intervention group overwhelmingly cited our new exercises as the most effective preparation for the lab 10 activities on phylogenetic tree construction and analysis (79% of responses; Figure 4). Of this group, approximately one-third of responses specified that in-lab collaborative work, and discussion time in particular, was helpful for them. When asked in what ways they felt unprepared for the phylogenetics lab, more than 60% of student responses indicated that they felt prepared or had problems with the format of the lab itself and not with the skills required to complete the lab (Figure 5). Some student responses (13%) indicated a desire for more practice with between-tree comparisons, while 11% of responses indicated a desire for more focus on the terminology used in the phylogenetics lab. Only a few students (6% of responses) indicated that they felt insufficiently prepared to construct phylogenetic trees.

DISCUSSION

Phylogenetic analysis and other evolutionary topics have proven to be perennially difficult concepts for undergrad-

uates, who maintain persistent misconceptions and incorrect reasoning (Anderson *et al.*, 2002; Baum *et al.*, 2005; Shtulman, 2006; Meir *et al.*, 2007; Perry *et al.*, 2008; Nehm and Ha, 2011). We expected that repeated exposure to phylogenetic concepts through increasingly difficult practice within an upper-level plant biology course would be effective for developing retention of knowledge, deep understanding of problems, and transfer of skills (Ericsson *et al.*, 1993; NRC, 2000; Billing, 2007; Larsen *et al.*, 2009; Chi and VanLehn, 2012). We implemented a deliberate practice instructional approach (Ericsson *et al.*, 1993), using a series of exercises that encouraged the incremental development of skills and content knowledge associated with phylogenetic analysis.

Our deliberate practice approach was effective in developing our students’ “tree-thinking” abilities. Final exam scores for phylogenetic questions significantly increased, while performance on nonphylogenetic questions remained unchanged, thereby improving our students’ overall performance in the course. In addition, students significantly improved in five of the 10 pre- and postconcept assessment questions. Finally, students found that our materials sufficiently prepared them to apply their tree-thinking skills in a culminating laboratory activity. Recognizing that students struggle to read traits from a phylogenetic tree, deduce ancestral traits, and reconstruct phylogenetic trees (Meir *et al.*, 2007; Perry *et al.*, 2008), our incremental approach addressed three categories of tree-thinking skills: interpreting relationships within trees, comparing among trees, and constructing trees.

Student Skills in Tree Interpretation

Students who experienced our tree-thinking assignments were better able to interpret relationships within trees. In contrast to students in the comparison group, students in the intervention group were better able to locate and identify

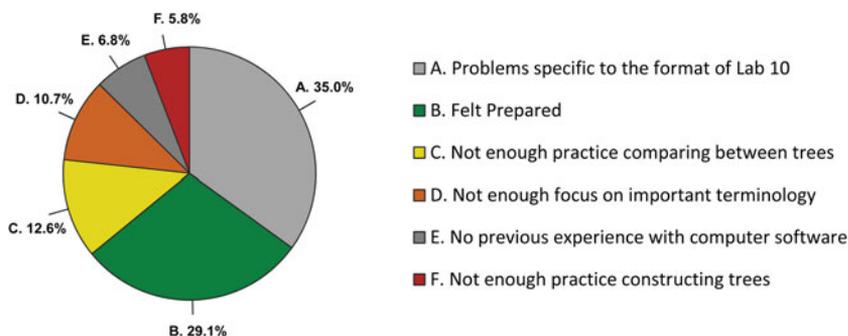


Figure 5. Summary of student responses to the question: “In what ways did you feel unprepared for the phylogenetic exercises of lab 10?” ($n = 103$; combined 2010 and 2011).

a source of homoplasy by discerning between the character traits mapped on a tree (Figure 2, Q.20b). These students also were more likely to correctly identify the extant outgroup of a tree (Figure 2, Q.17a). Our new assignments reinforced both of these skills through repeated exposure to constructing trees that contained extant outgroups and instances of homoplasy. In the pre- and postconcept assessment, students showed significant improvements in their ability to perform within-tree comparisons on three items (Table 3, items 1, 3, and 6). A fourth item that addressed within-tree comparisons did not show a significant improvement (Table 3, item 2). However, similar to the results reported by Meir *et al.* (2007), the majority of the students entering our class could already correctly map the direction of time on a tree, greatly reducing the effective sample size of the normalized change metric. Both the final exam scores and the pre- and postconcept assessment strongly suggest that our new assignments significantly improved our students' ability to interpret relationships within a tree.

Student Skills in Tree Comparison

The ability to compare between phylogenetic trees represents a substantial component of the lab 10 exercises in our course. In this lab, students are asked to evaluate the similarities and differences between two competing trees that were constructed from independently derived character traits. Despite the emphasis on this skill in lab 10, only one pre-existing final exam question addressed between-tree comparisons. Students in the intervention group performed significantly better on this question, which requires their understanding of homoplasy to be transferred to a novel situation that involves comparing character state changes on six separate trees (Figure 2, Q.2a). One question in the pre- and postconcept assessment addressed between-tree comparisons, and students did not show significant improvements on this question (Table 3, item 4). However, in contrast to the results reported by Meir *et al.* (2007), the vast majority of our students, in both the pre- and postconcept assessments, understood that rotating the branches of a tree at the nodes did not affect the evolutionary relationships of the taxa located at the branch tips, resulting in a small effective sample size for the normalized change metric.

While our assessments provide some support that our students' skills in tree comparison improved as a result of our new assignments, both the assignments and the assessments could more directly address this skill in the future. While the most common response to the student feedback survey was that students felt prepared for the exercises of lab 10, the second most common response was that students thought they needed more preparation comparing between trees. Our new assignments did engage students in between-tree comparisons to some extent, as the tree construction exercises of the final three assignments were intentionally designed to have multiple most-parsimonious answers. The possibility of obtaining several correct answers created opportunities for in-lab discussions about the methods for comparing between competing phylogenetic trees, such as calculating consistency indices. However, the assignments themselves could have included more structured opportunities for our students to compare between two or more trees.

Student Skills in Tree Construction

Student performance in tree construction tasks improved between the pre- and postconcept assessments on two of the four tree construction items. However, there was no significant improvement in our students' ability to construct the most parsimonious branching pattern (Table 3, item 7a). This result was unexpected because correctly constructing the most parsimonious tree, given a set of taxa and traits, forms a primary component of our exercises. Furthermore, high scores obtained in the challenging tree construction question from the second laboratory exam, along with positive reviews in the student feedback surveys, indicate that our assignments did a better job at improving our students' tree construction skills than is conveyed by the pre- and postconcept assessment data. One possible explanation for this result involves the simplicity of the concept assessment question. Although students in the comparison group struggled with constructing the relatively complex trees in lab 10 prior to the implementation of our exercises, the majority of students in the intervention group (72.7%) obtained perfect scores on this component of the tree construction problem in the preconcept assessment. The concept assessment question uses a small number of taxa and traits, represented pictorially, and results in a simple tree with only three nodes and no instances of complicating factors, such as homoplasy or a polytomy. In contrast, the trees constructed in our new assignments and the lab 10 exercises contain six to 11 nodes and multiple instances of homoplasy and/or polytomies. Interestingly, 18 students in the postconcept assessment unnecessarily included instances of homoplasy in their answer to the tree construction question, thus resulting in a tree that was not the most parsimonious, while only one student made this mistake in the preconcept assessment. In other words, students received lower scores on the postconcept assessment because they provided answers that were more complicated than necessary.

Students' ability to correctly place an extinct common ancestor decreased between the pre- and postconcept assessments. In the postconcept assessment, students frequently drew the extinct common ancestor as if it were an extant outgroup. The intervention assignments and lab 10 exercises presented only extant outgroups to the students, and the familiarity students gained with extant outgroups between the pre- and postconcept assessments may explain why so many students would mistakenly place an extinct ancestor at the tip of the tree. The decrease in student performance on this component of tree construction appears to represent a mismatch between our assessment and our teaching materials, a problem that could be easily remedied by modifying the tree construction question in the pre- and postconcept assessment.

CONCLUSION

Our application of the deliberate practice approach to improve students' tree-thinking skills responds to several current issues in biology education. First, our exercises addressed common difficulties that students experience with evolutionary concepts and taught students to think like scientists in the creation of phylogenetic trees. Furthermore, our intervention has the potential to increase undergraduate

retention in the sciences because it focuses on improving the experiences of undergraduates in a science lab course (O'Neal *et al.*, 2007). Finally, our skill-focused intervention responds to criticisms of large science courses that are teacher centered and content heavy, with exams that represent independent topics and require superficial understanding of content and little demonstration of skills (Handelsman *et al.*, 2004; Alberts, 2005; Wood, 2009). Specifically, our study contributes further support for student-centered approaches that encourage learners to connect new knowledge to prior knowledge through conversation with colleagues (Lord, 1997) and practice with increasingly complex problem-solving tasks (NRC, 2000; Billing, 2007; Chi and VanLehn, 2012).

Overall, our findings confirm that a deliberate practice approach improves the ability of college students to understand challenging concepts and acquire difficult skills. Part or all of our new teaching materials should be adaptable to a range of introductory through advanced biology courses that include a component of phylogenetics, such as evolution, systematics, or organismal biology courses. Our assignments have been implemented in the classroom by six different GTAs, four of whom were not involved in creating the assignments, and none of whom specializes in phylogenetics as part of their doctoral studies. We recognize that most biology laboratory experiences are tightly booked with a variety of activities, and we designed our assignments to require little additional class time or instructor resources. However, an important component of class time was group discussion of the assignments. These in-class discussions gave students a solid grounding in the terminology underlying the concepts and contributed to our students' success on tree construction tasks through active participation. Given our positive results with this method, we recommend a deliberate practice approach to teaching difficult concepts in the life sciences.

ACKNOWLEDGMENTS

The authors conducted this project with funding support from the Teagle Foundation through the Indiana University Collegium on Inquiry in Action (www.iub.edu/~collegia). We especially thank collegium members Mimi Zolan, Melissa Gresalfi, and Tyler Christensen, who provided advice on the design of the intervention and assessment materials. We thank Eli Meir, whose misconception pre- and posttest provided inspiration for our own pre- and postconcept assessment. We also thank the faculty instructor for B300, Eric Knox, and the GTAs who taught in B300 during the duration of the project (W. Beaulieu, M. Bogonovich, B. Mattingly, M. Midgley, S. Shannon, and B. Swedo), as well as our B300 students from 2007 to 2011.

REFERENCES

- Alberts B (2005). A wakeup call for science faculty. *Cell* 123, 739–741.
- Anderson DL, Fisher KM, Norman GJ (2002). Development and evaluation of the conceptual inventory of natural selection. *J Res Sci Teach* 39, 952–978.
- Anderson LW, Krathwohl DR, Airasian PW, Cruikshank KA, Mayer RE, Pintrich PR, Raths J, Wittrock MC (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, abridged ed., New York: Longman.
- Bauer-Dantoin A (2009). The evolution of scientific teaching within the biological sciences. In: *Exploring Signature Pedagogies: Approaches to Teaching Disciplinary Habits of Mind*, ed. RAR Gurung, NL Chick, and A Haynie, Sterling, VA: Stylus, 224–243.
- Baum DA, Offner S (2008). Phylogenies and tree-thinking. *Am Biol Teach* 70, 222–229.
- Baum DA, Smith SD, Donovan SSS (2005). The tree-thinking challenge. *Science* 310, 979–980.
- Berkman MB, Plutzer E (2011). Defeating creationism in the courtroom, but not in the classroom. *Science* 331, 404–405.
- Billing D (2007). Teaching for transfer of core/key skills in higher education: cognitive skills. *High Educ* 53, 483–516.
- Burks RL, Boles LC (2007). Evolution of the chocolate bar: a creative approach to teaching phylogenetic relationships within evolutionary biology. *Am Biol Teach* 69, 229–237.
- Cepeda NJ, Coburn N, Rohrer D, Wixted JT, Mozer MC, Pashler H (2009). Optimizing distributed practice theoretical analysis and practical implications. *Exp Psychol* 56, 236–246.
- Chi MTH, VanLehn KA (2012). Seeing deep structure from the interactions of surface features. *Educ Psychol* 47, 177–188.
- Chick NL, Haynie A, Gurung RAR (2009). From generic to signature pedagogies: teaching disciplinary understandings. In: *Exploring Signature Pedagogies: Approaches to Teaching Disciplinary Habits of Mind*, ed. RAR Gurung, NL Chick, and A Haynie, Sterling, VA: Stylus, 1–16.
- Cohen J (1960). A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20, 37–46.
- Crippen KJ, Brooks DW (2009). Applying cognitive theory to chemistry instruction: the case for worked examples. *Chem Educ Res Pract* 10, 35–41.
- Crowe A, Dirks C, Wenderoth MP (2008). Biology in Bloom: implementing Bloom's taxonomy to enhance student learning in biology. *CBE Life Sci Educ* 7, 368–381.
- Duncan RG, Hmelo-Silver CE (2009). Learning progressions: aligning curriculum, instruction, and assessment. *J Res Sci Teach* 46, 606–609.
- Ericsson KA, Krampe RT, Tesch-Römer C (1993). The role of deliberate practice in the acquisition of expert performance. *Psychol Rev* 100, 363.
- Freeman S, Haak D, Wenderoth AP (2011). Increased course structure improves performance in introductory biology. *CBE Life Sci Educ* 10, 175–186.
- Goldsmith DW (2003). The great clade race—presenting cladistic thinking to biology majors and general science students. *Am Biol Teach* 65, 679–682.
- Haak DC, HilleRisLambers J, Pitre E, Freeman S (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science* 332, 1213–1216.
- Hackling M, Garnett P (1992). Expert-novice differences in science investigation. *Res Sci Educ* 22, 170–177.
- Hake RR (1998). Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *Am J Phys* 66, 64–74.
- Halverson KL, Pires CJ, Abell SK (2011). Exploring the complexity of tree thinking expertise in an undergraduate systematics course. *Sci Educ* 95, 794–823.
- Handelsman J *et al.* (2004). Scientific teaching. *Science* 304, 521–522.
- Kennedy CH (2005). *Single-Case Designs for Educational Research*, Boston: Allyn & Bacon.
- Knight JK, Wood WB (2005). Teaching more by lecturing less. *Cell Biol Educ* 4, 298–310.
- Krathwohl DR (2002). A revision of Bloom's taxonomy: an overview. *Theory Pract* 41, 212–218.

- Larsen DP, Butler AC, Roediger HL (2009). Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Med Educ* 43, 1174–1181.
- Lents NH, Cifuentes OE, Carpi A (2010). Teaching the process of molecular phylogeny and systematics: a multi-part inquiry-based exercise. *CBE Life Sci Educ* 9, 513–523.
- Lord TR (1997). A comparison between traditional and constructivist teaching in college biology. *Innov High Educ* 21, 197–216.
- Marx JD, Cummings K (2007). Normalized change. *Am J Phys* 75, 87–91.
- Meir E, Perry J, Herron JC, Kingsolver J (2007). College students' misconceptions about evolutionary trees. *Am Biol Teach* 69, 71–76.
- Middendorf J, Pace D (2004). Decoding the disciplines: a model for helping students learn disciplinary ways of thinking. *New Dir Teach Learn* 98, 1–12.
- Mohan L, Chen J, Anderson CW (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *J Res Sci Teach* 46, 675–698.
- National Research Council (2000). *How People Learn: Brain, Mind, Experience, and School*, expanded ed., Washington, DC: National Academies Press.
- National Research Council (2007). *Learning progressions*. In: *Taking Science to School: Learning and Teaching Science in Grades K–8*, Washington, DC: National Academies Press, 213–250.
- Nehm RH, Ha MS (2011). Item feature effects in evolution assessment. *J Res Sci Teach* 48, 237–256.
- O'Neal C, Wright M, Cook C, Perorazio T, Purkiss J (2007). The impact of teaching assistants on student retention in the sciences: lessons for TA training. *J Coll Sci Teach* 36, 24–29.
- Perry J, Meir E, Herron JC, Maruca S, Stal D (2008). Evaluating two approaches to helping college students understand evolutionary trees through diagramming tasks. *CBE Life Sci Educ* 7, 193–201.
- Plant EA, Ericsson KA, Hill L, Asberg K (2005). Why study time does not predict grade point average across college students: implications of deliberate practice for academic performance. *Contemp Educ Psychol* 30, 96–116.
- Shtulman A (2006). Qualitative differences between naïve and scientific theories of evolution. *Cogn Psychol* 52, 170–194.
- Shtulman A, Valcarcel J (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition* 124, 209–215.
- Simpson MG (2010). *Plant Systematics*, 2nd ed., Amsterdam: Elsevier Academic.
- Tanner K, Allen D (2005). Approaches to biology teaching and learning: understanding the wrong answers—teaching toward conceptual change. *Cell Biol Educ* 4, 112–117.
- van Gog T, Ericsson KA, Rikers RM, Paas F (2005). Instructional design for advanced learners: establishing connections between the theoretical frameworks of cognitive load and deliberate practice. *Educ Tech Res Dev* 53, 73–81.
- Wiggins GP, McTighe J (2005). *Understanding by Design*, 2nd ed., Alexandria, VA: Association for Supervision and Curriculum Development.
- Wood WB (2009). Revising the AP biology curriculum. *Science* 325, 1627–1628.
- Yu DD (2011). How much do study habits, skills, and attitudes affect student performance in introductory college accounting courses? *New Horizons Educ* 59, 1–15.