

Feature Current Insights

Recent Research in Science Teaching and Learning

Deborah Allen

Department of Biological Sciences, University of Delaware, Newark, DE 19716

This feature is designed to point *CBE—Life Sciences Education* readers to current articles of interest in life sciences education, as well as to more general and noteworthy publications in education research. URLs are provided for the abstracts or full text of articles. For articles listed as “Abstract available,” full text may be accessible at the indicated URL for readers whose institutions subscribe to the corresponding journal.

1. White B, Stains M, Escriu-Sune M, Medaglia E, Rostamjad L, Chinn C, Sevian H (2011). A novel instrument for assessing students’ critical thinking abilities. *J Coll Sci Teach* 40, 102–107.

[Minimal abstract available: www.nsta.org/publications/browse_journals.aspx?action=issue&id=10.2505/3/jcst11_040_05]

This article reports on the development and initial use of a relatively simple instrument designed to assess critical-thinking skills in the context of science concepts and processes. To situate their work, the authors cite the Delphi Report’s (Facione, 1990) definition of critical thinking as “purposeful, self-regulatory judgment that results in interpretation, analysis, evaluation, and inference as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based.” They have placed this broader consideration of what is meant by critical thinking in the more specific context of science by designing an Assessment of Critical Thinking Ability (ACTA) that asks the “test-taker” to draw conclusions about evidence collected in scientific studies, particularly in instances in which the evidence from multiple studies supports different, perhaps conflicting, conclusions. The survey is designed to assess three major subsets of critical-thinking competencies needed to reason effectively when presented

with multiple lines of evidence, as follows: Ability 1: drawing a unified conclusion from conflicting evidence; Ability 2: designing experiments that could resolve “murky” areas of previously conducted studies; and Ability 3: formulating alternatives to the original conclusions about a study.

The instrument was constructed to assess these abilities by presenting three partially conflicting and related case studies, each of which describes the methods and data collected from research studies about a similar topic. In the example provided, the cases are about the causes of the disease pellagra. The subjects taking the instrument are first asked to evaluate each study individually, then to draw a conclusion (in the example, about which disease cause a study supports) from the evidence presented in each. They are then required to weigh each set of evidence against the others, to determine which best supports a given conclusion (in the example, which cause is the most likely one for pellagra), and to report on their confidence in this conclusion, as well as on the likelihood that someone else would be able to draw a different, valid conclusion based on the same evidence. The authors provide information about the construct validity of the paper (including its readability), along with a four-tiered scoring rubric that reflects increasing levels of competence in each of the three critical thinking–ability subsets.

The article also reports on data obtained from administration of the instrument to four different groups of participants at successively higher academic levels: 106 students in a first-year biology course, 47 senior-level science majors, 19 biology and chemistry graduate students, and 13 postdoctoral fellows in biology and chemistry. The results in general show a statistically significant positive relationship between the participants’ years of exposure to academic science and their overall ACTA scores. However, more finely grained analysis of relative scores in each of the three ability areas indicate there were dissimilar patterns of difference among the three abilities. That is, while graduate students scored significantly above the senior-level undergraduates in Ability 2, their scores were not significantly different in Abilities 1 and 3. The authors attribute this dissimilar pattern to different levels of difficulty in achieving competency in the three areas. For example, more students at all levels fell into the highest rubric category for Ability 1, possibly because it may be easiest to master. In contrast, very few at all academic levels scored at the highest rubric category for Ability 2.

DOI: 10.1187/cbe.11-09-0086

Address correspondence to: Deborah Allen (deallen@udel.edu).

© 2011 D. Allen. *CBE—Life Sciences Education* © 2011 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.

The authors conclude by discussing the implications of the instrument and the initial study results for the teaching community—the potential utility of the ACTA in development and assessment of curricula that would hone students' critical-thinking skills.

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

[Abstract available: <http://onlinelibrary.wiley.com/doi/10.1002/sce.20436/abstract>]

In this study, the authors explore how students interpret phylogenetic trees, a form of external representation that conveys essential information about evolutionary histories or inferred relationships among taxa. The work builds on prior studies documenting some of the difficulties students have with interpreting phylogenetic trees by attempting to determine the common patterns of reasoning and problem-solving approaches that might underlie students' misconceptions about what information trees convey, how trees are constructed, and how the information presented by trees can be used to inform the resolution of systematics problems.

The study took place in the context of an upper-level, undergraduate, plant systematics course that included building students' tree-thinking expertise as one of its objectives. Thirty-five students agreed to participate in the study through class activities alone, with nine of the 35 (the key informants) agreeing to additional participation in an interview process. Class activities used for data collection included a two-tiered pre/posttest that was administered to all participants and classroom observations. The authors observed students working with phylogenetic trees and collected additional information about instructional approaches, student questions and comments, etc., during each class session. Additional data were collected midsemester through 90-min, semistructured interviews with the nine key informants.

The pre- and posttest responses and interview transcripts were the key sources of data used for analysis. Responses were coded according to the patterns that emerged in the data about the types of representation-based problem-solving approaches and reasoning patterns used by the students, as evidenced by how they justified their responses. With respect to representation-based problem-solving ability, the authors identified and classified three categories of ability that reflected increasing levels of expertise from novice, through intermediate, to expert level. Most of the students fell into the intermediate category, characterized as one in which inability to piece together all of the interconnected concepts related to evolutionary biology, tree thinking, and systematics hindered their ability to fully interpret and use the information in phylogenetic trees.

The authors identified students' reasoning processes as falling within one of the following eight types, characterized according to degree of resemblance to the processes exhibited by experts (but not intended to represent a linear developmental path to expertise): 1) extraneous, 2) ecological, 3) morphological, 4) branch influenced, 5) tree-shape influenced, 6) node influenced, 7) limited phylogenetic, and 8) phylogenetic. For each category, the authors provide further descriptions of its distinct features, and of the lines of evidence from tests and interviews that formed the basis for formulating

each category and assigning students to it. For the ecological category, for example, students erroneously factored ecological information, such as organisms' typical habitats, to identify species and their relatedness (more specifically, e.g., concluding that seals and whales must be more closely related than either is to horses, because both are aquatic). Node-influenced reasoning, a common intermediate pattern, was characterized by a focus on use of the number and location of the nodes depicted on a tree, but misuse of the information that nodes are intended to convey. A student's use of a particular reasoning type was not fixed, but could vary with the nature of the representation-based, problem-solving task with which he or she was confronted.

The authors conclude with a statement about the instructional implications of their findings. They posit that the findings reveal that a common source of novice students' misinterpretation of the information conveyed by trees stems from their inclination to apply extraneous, nonphylogenetic information. Thus, a more gradual and explicit instructional focus on the development of foundational aspects of evolutionary reasoning before introduction of the use of complex graphical representations, such as phylogenetic trees in systematics problem-solving, might be required to help students effectively meet the "tree thinking challenge" (Baum *et al.*, 2005). Additionally, the authors suggest that their study findings highlight the value of qualitative assessments in shedding light on complexities of thinking not revealed by the more commonly used multiple-choice survey instruments. The article is accompanied by appendices that include sample items used in the tests and interviews.

3. Partin ML, Haney JJ, Worch EA, Underwood EM, Nurnberger-Haag JA, Scheuermann A, Midden WR (2011). Yes I can: the contributions of motivation and attitudes on course performance among biology nonmajors. *J Coll Sci Teach* 40, 86–95.

[Minimal abstract available: www.nsta.org/publications/browse_journals.aspx?action=issue&id=10.2505/3/jcst11_040_06]

The authors report on the results of administering a trio of survey instruments in a nonmajors biology course: the Biology Attitude Scale (BAS), the Mathematics Attitude Scale (MAS), and a portion of the Motivated Strategies for Learning Questionnaire (MSLQ). They set the stage for the reader to be able to follow their interpretation of the findings by describing some of the constructs the surveys were designed to measure and their complex interplays, including intrinsic and extrinsic motivation, beliefs and attitudes about science, and self-efficacy (put simply, a person's belief about their ability to succeed in a given situation; Bandura, 1977); the factors thought to influence these constructs; and their relationships to student achievement. The authors used the instruments to explore questions related to whether these variables could predict course performance (and whether some were better predictors than others).

The administration of the surveys was conducted in four sections of a one-semester, nonmajors biology course, taught in a lecture format by different instructors. The authors provide demographic information for the 318-student population of study participants (63% response rate for surveys). The three survey instruments were administered in the last

2 wk of the semester; data about final grades (converted to a z-score to account for any differences in grading practices between the sections) were collected, along with the survey data, as a measure of course performance. The authors provide additional pertinent information about the survey instruments, including their sources, item numbers subscales (for the MSLQ), and reliability.

Analysis of survey data revealed that the most significant predictors of course performance were self-efficacy, attitudes about math, and test anxiety, with self-efficacy being the strongest predictor of the three. Given the importance of self-efficacy as a predictor, the authors further analyzed the data (multiple regression analysis) to determine whether any independent variables contributed to students' sense of self-efficacy. The analysis indicated that six variables could be factored into the model, with varying effect sizes—intrinsic and extrinsic goal orientation, perceived value of the task (the activities of the course), test anxiety, control of learning beliefs (belief that effort to learn will produce positive outcomes [an MSLQ subscore]), and attitudes about biology and math. Of these, attitudes about biology and control of learning beliefs were the strongest predictors of self-efficacy.

The results reported for this study are not widely discrepant from previous studies conducted in the context of students in college-level science courses. What makes this work more distinct is that the authors conclude by addressing the implications of their work in a substantive way. They present the basic features of a research-based, C-5 Model that provides strategies for creating a biology classroom learning environment designed to enhance student motivation.

4. Ebert-May D, Derting TL, Hodder J, Momsen JL, Long TM, Jardeleza SE (2011). What we say is not what we do: effective evaluation of faculty professional development programs. *BioScience* 61, 550–558.

[Abstract available: www.jstor.org/pss/10.1525/bio.2011.61.7.9]

The effectiveness of faculty professional development (PD) workshops and programs aimed at assisting higher education faculty in implementing new teaching and learning paradigms is commonly assessed using surveys that prompt faculty self-reports of satisfaction with the overall organization, the quality of presentation of specific sessions, and what they have learned and intend to implement in their classrooms. Less commonly, there are follow-up surveys that ask faculty to report on the instructional strategies they actually do begin to use or refine as the result of participation in the PD program. The authors of this study report on an important additional step: the findings from independent assessment of the actual classroom practices of the program participants as a supplement to the information provided by the survey self-reports.

The study entailed collection and analysis of data obtained from participants in two programs with similar overarching goals related to fostering use of and knowledge about active learning pedagogies: the Faculty Institutes for Reforming Science Teaching (FIRST II) and the National Academies Summer Institute on Undergraduate Education in Biology (SI). The FIRST II program participants attended sessions on 6–12 d spread out over 3 yr, while the SI program brought in a new cohort of faculty each year for a 5-d summer experience and a 2-d follow-up in the midst of the ensuing year.

There were two main data sources for analysis: 1) baseline, midprogram, midcourse, and final surveys that explored aspects of participants' pedagogical knowledge and experiences with active-learning pedagogies; and 2) one or more videotapes of participants' actual teaching during class sessions. The final survey was conducted at the end of the PD program, as was the videotape analysis (tapes were collected during the 2 yr following the PD workshops), which used the Reformed Teaching Observation Protocol (RTOP). Two hundred eleven participants provided survey data, and a subset of 77 participants participated in the classroom observation (RTOP) study.

Use of the RTOP instrument's associated rating system allowed for assignment of subscale scores to the classroom videos by the observers; each tape was independently rated by two independent observers trained in the protocol. For those unfamiliar with it, the RTOP subscale scores fall under five areas (lesson design and implementation, propositional knowledge, procedural knowledge, communicative interactions, and teacher–student relationships). The sum of the subscale scores assigned by the rater (total score) falls into one of five categories that reflect the relative degree of implementation of active learning. The first two categories are characteristic of classrooms in which the teaching practices are relatively teacher-centered (use of lecture alone, or lecture with some demonstrations), and the fourth and fifth are assigned to classrooms showing the hallmarks of student engagement in various facets of inquiry. A score for a typical college-level “reformed” course that blends teacher- and student-centered activities would fall under the third, intermediate category.

In addition to this standard RTOP analysis, the authors performed statistical comparisons between the sets of survey data collected at different times. They used the survey data and information about faculty demographics to formulate 10 variables that were potential predictors of teaching practices (e.g., class size, instructor confidence level, proportion of workload assigned to teaching, challenges encountered), then used a general linear models effect selection analysis to identify the utility of these variables as predictors of the total and subscale RTOP scores. Initial analyses revealed there was a legitimate basis for combining the data from the FIRST II and SI program faculty for reporting purposes.

The analyses revealed that the majority of faculty who had participated in one of the two PD programs showed gains in knowledge about various aspects of reformed teaching, and reported that they had implemented specific inquiry-based and learner-centered teaching practices after their participation. However, the data resulting from classroom observations were not well aligned with faculty perceptions: RTOP analyses revealed that the majority of faculty (roughly 75%) actually implemented relatively teacher-centered practices, with average total RTOP scores that fell into category I or II. There were no major shifts in these practices with greater elapsed time post-PD. Four of the potential 10 predictors of the RTOP score accounted for 20% of the variation and were entered into the prediction model: number of years of undergraduate teaching experience, class size, the proportion of time assigned to teaching (all three were inversely related to the RTOP score), and the instructor's direct experience with active learning (greater extent of experience with active learning associated with higher scores). More conservative selection criteria schemes allowed for inclusion of only the first of

these two variables into the model. Interestingly, about 80% of the observed variation remained unaccounted for by any of the linear regression models.

The article concludes by placing the discrepancy between data collected by self-report (surveys) and those collected by independent observers (RTOP analysis of videos) in the context of other studies reported in the literature—that is, this disconnect between perception and reality in the arena of professional development for teachers is not unusual. The authors discuss possible explanations for why the variable “years of teaching” was negatively related to RTOP scores, and reinforce previous recommendations that PD is most likely to be effective early in a faculty member’s career. Finally, they offer some suggestions for re-envisioning faculty PD programs that mirror some of the “community of practice” and apprenticeship models that are becoming increasingly widespread in the K–12 setting.

I invite readers to suggest current themes or articles of interest in life science education, as well as influential papers published in the more distant past or in the broader field of education research, to be featured in *Current Insights*. Please send any suggestions to Deborah Allen (deallen@udel.edu).

REFERENCES

- Bandura A (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev* 84, 191–215.
- Baum DA, Smith SD, Donovan SS (2005). The tree-thinking challenge. *Science* 310, 979–980.
- Facione PA (1990). *Critical thinking: a statement of consensus for purposes of educational assessment and instruction*, Millbrae, CA: California Academic Press.