

Article

Using Assessments to Investigate and Compare the Nature of Learning in Undergraduate Science Courses

Jennifer Momsen,* Erika Offerdahl,[†] Mila Kryjevskaja,[‡] Lisa Montplaisir,*
Elizabeth Anderson,* and Nate Grosz[‡]

*Department of Biological Sciences, [†]Department of Chemistry & Biochemistry, and [‡]Department of Physics,
North Dakota State University, Fargo, ND 58108

Submitted September 1, 2012; Revised February 14, 2013; Accepted February 14, 2013
Monitoring Editor: Eric Brewé

Assessments and student expectations can drive learning: students selectively study and learn the content and skills they believe critical to passing an exam in a given subject. Evaluating the nature of assessments in undergraduate science education can, therefore, provide substantial insight into student learning. We characterized and compared the cognitive skills routinely assessed by introductory biology and calculus-based physics sequences, using the cognitive domain of Bloom's taxonomy of educational objectives. Our results indicate that both introductory sequences overwhelmingly assess lower-order cognitive skills (e.g., knowledge recall, algorithmic problem solving), but the distribution of items across cognitive skill levels differs between introductory biology and physics, which reflects and may even reinforce student perceptions typical of those courses: biology is memorization, and physics is solving problems. We also probed the relationship between level of difficulty of exam questions, as measured by student performance and cognitive skill level as measured by Bloom's taxonomy. Our analyses of both disciplines do not indicate the presence of a strong relationship. Thus, regardless of discipline, more cognitively demanding tasks do not necessarily equate to increased difficulty. We recognize the limitations associated with this approach; however, we believe this research underscores the utility of evaluating the nature of our assessments.

INTRODUCTION

Undergraduate science classrooms are complex systems made up of multiple, interacting components with the ultimate function of impacting student learning. As such, attempts at understanding student learning must characterize the individual components of the system (e.g., instructor, student, curriculum), while simultaneously elucidating the nature of the interactions between those components. Scholars from fields ranging from education to cognitive

psychology have, through theoretical and empirical work, identified several influential components of the system that provide a foundation from which discipline-based education researchers can investigate student learning in the context of science. Emergent from this literature are two salient ideas. The first is that student learning is mediated by an array of factors, including students' conceptions of and approaches to learning (Prosser and Trigwell, 1999), motivation and ability to self-regulate learning (Zimmerman *et al.*, 1992), and perceptions of the learning environment (Trigwell and Prosser, 1991; Gijbels *et al.*, 2008). The second is that changes in the learning environment can affect any of these factors and, as a result, the learning outcomes achieved (Prosser and Trigwell, 1999; Micari and Light, 2009).

More than a means of assigning a grade, assessments, both formative and summative, represent one contextual element of a student's learning environment, and therefore may significantly shape students' learning through their perceptions of the content and cognitive skills needed to do well (Hammer *et al.*, 2005; Hall *et al.*, 2011). Students may selectively or strategically study in a manner that best matches their expectations of the assessment (Entwistle and Entwistle,

DOI: 10.1187/cbe.12-08-0130

Address correspondence to: Jennifer Momsen (jennifer.momsen@ndsu.edu).

© 2013 J. Momsen *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.

1992). For example, if an exam is likely to be filled with multiple-choice questions primarily testing recall and basic comprehension, students might choose a surface-learning approach, spending study time memorizing facts, processes, and formulas. By extension, students might transfer the learning approach to ideas about the nature of knowledge, viewing the discipline as factual, disparate, and disconnected.

Classroom assessment has the potential to powerfully influence student learning by contributing directly to the so-called “hidden curriculum,” an informal, unarticulated, and often unintended curriculum (Snyder, 1973; Entwistle, 1991) affecting what, how, when, and how much students learn (Crooks, 1988; Scouller and Prosser, 1994; Scouller, 1998). Thus, our study is informed by the premise that assessment is a key contextual element of the learning environment that can both shape and reinforce students’ ideas about the nature of knowledge and learning, as well as the skills and disciplinary practices of scientists. Such ideas can impact student learning within each discipline and may also affect students’ abilities to transfer skills and knowledge across disciplines.

This study builds on prior work examining 100- and 200-level undergraduate biology courses for both majors and non-majors to focus more specifically on the nature of assessment in introductory courses that serve as entry points for biology and physics majors. We used Bloom’s taxonomy to characterize the cognitive skills regularly assessed in two introductory science sequences for discipline majors: introductory biology and introductory calculus-based physics. Specifically, we characterized and compared 1) cognitive skills of assessments and 2) the relationship between cognitive skill level and student performance. The data presented provide preliminary insights into the complex set of factors affecting student learning both in a single course and across disciplines.

Mediating Factors of Student Learning Outcomes

Students’ approaches to learning have long been investigated as a possible mediating factor of the learning achieved by students. While a number of approaches to learning have been operationalized in the literature, three emerge as the most common: surface, deep, and strategic/achieving (Table 1; e.g., Marton and Säljö, 1976; Biggs, 1987; Entwistle *et al.*, 2000). Students who adopt surface approaches to learning are more likely to memorize information, interact passively with material, and seldom engage in reflection, and are less likely to make connections between ideas or recognize emergent patterns. In contrast, a deep approach to learning is commonly associated with efforts toward obtaining conceptual understanding, critical analysis of information, actively engaging with material, and looking for relationships between ideas. The adoption of a third approach, referred to as either strategic or achieving, is heavily influenced by the desire to succeed on learning tasks and consists of characteristics of both deep and surface approaches. Students’ approaches to learning are not commonly viewed as stable traits; for example, a student may take a surface approach to learning in a mathematics course, while adopting a deep approach in history.

Similarly, research on students’ epistemological beliefs further reveals how students’ ideas about knowledge and its construction may impact learning. At the college level, Hall *et al.* (2011) focused on the discipline-specific and context-dependent nature of these ideas by examining what they call

Table 1. Common learning approaches^a

Learning approach	Description
Deep approach	Intention to understand for oneself Interacting vigorously and critically with the content Relating ideas to previous knowledge and experience Integrating components through organizing principles Relating evidence to conclusions Examining the logic of the argument
Strategic/achieving	Focused on success over learning Assessment oriented May use both deep and surface approaches
Surface approach	Intention simply to reproduce parts of the content Accepting ideas and information passively Concentrating only on assessment requirements Not reflecting on purpose or strategies Memorizing facts and procedures Failing to distinguish guiding principles or patterns

^aBased on Entwistle and Entwistle (1992) and Marton and Säljö (1976).

epistemological expectations, students’ beliefs about the nature of the knowledge being learned and what they need to do in order to learn in a particular course. Within the context of an undergraduate introductory biology sequence, they found that students tend to 1) characterize the nature of biological knowledge as facts provided by an authoritative source that do not need to be understood within any larger context and 2) believe that learning in introductory biology courses does not require math or physics. But they also found that epistemological expectations are dynamic; they can shift depending on how the student interprets contextual cues in the learning environment. Work by Hammer *et al.* (2005) in physics further substantiates the role of context in shaping student ideas about effectiveness in learning through their use of *epistemological frames*. They define “frame” as a set of expectations held by an individual with regard to a given situation that ultimately shapes what he or she pays attention to and how he or she acts. With respect to knowledge construction and learning, epistemological frames affect both the cognitive resources accessed during the learning process and student thinking about constructing new knowledge. Epistemological frames answer the question “How should one approach knowledge?” Though nuances distinguish the different approaches to investigating student epistemologies, there is growing evidence that, similar to students’ approaches to learning, students’ beliefs are dynamic, shaped by context, and heavily influence the learning process.

Student learning is also impacted by the degree to which students are successful at self-regulating their learning. Students who are effective self-regulators are able to 1) direct their learning by establishing challenging personal goals (Schunk, 1990), 2) recognize and apply appropriate strategies for achieving those goals, and 3) exhibit well-developed

metacognitive abilities (Zimmerman *et al.*, 1992). Students' personal goals for learning can also be influenced by externally assigned goals (Locke and Latham, 1990). It is reasonable to predict, therefore, that course learning objectives, both implicit and explicit, are likely to have an impact on what and how students approach and engage in learning. Moreover, the achievement of course objectives by students is dependent, at least in part, on the extent to which students are able to employ the requisite metacognitive skills to interpret the task at hand and to apply the appropriate skills and approaches necessary to meet those objectives.

Assessing Assessment

Assessments are an integral component of the teaching and learning process, particularly when viewed as a contextual element of the learning environment to which students pay attention. As a result, the potential impact of assessment on student learning warrants a deeper investigation through the examination of assessment practices routinely used in science classrooms. Given the limited training of instructors in both K–12 and undergraduate classrooms in developing assessments that align with stated learning goals and instructional practices (Carter, 1984; Stiggins, 2001), it is not surprising to find assessments that 1) mirror how instructors themselves were taught (Lortie, 1975), 2) contain easy to assess content and concepts (Frederiksen, 1984; Crooks, 1988), or 3) reflect a “facts first” mentality, wherein an instructor believes students must first know basic facts, concepts, and processes before moving on to more complex cognitive tasks (Guo, 2008). Indeed, some instructors believe that questions requiring students to demonstrate complex cognitive skills may serve to confuse students, perpetuate misconceptions, and ultimately end in failure for students (Frederiksen, 1984; Doyle, 1986). Further, writing test items at higher cognitive levels is a known difficulty (Crooks, 1988). As a result, assessment of learning in science courses may not align with desired learning outcomes and may send a tacit, unintended message to students about the nature of the learning in which they should engage. This may be particularly true of introductory courses, which serve an increasing number of undergraduates who are diverse in their college preparation, academic abilities, career goals, and motivations to learn. In response to such diversity, these courses are often tasked with teaching scientific content and disciplinary practices alongside basic scientific literacy and learning skills.

Characterizing the cognitive skill levels of assessments can employ one of several taxonomies, including Webb's “depth of knowledge” (Webb, 1997), Bloom's taxonomy (Bloom, 1956), and even a revised Bloom's taxonomy that combines the cognitive domain taxonomy with the affective domain taxonomy (Anderson and Krathwohl, 2001; Krathwohl, 2002). We selected the original Bloom's taxonomy, *Taxonomy of Educational Objectives: The Cognitive Domain*, in part because it is widely used in biology education research, but also because it is readily accessible to scientists (Crowe *et al.*, 2008).

Bloom's taxonomy identifies six levels of cognitive skills that are routinely targeted by educational objectives (Table 2), which represent simple to complex cognitive skills and concrete to abstract thinking (Krathwohl, 2002). These cognitive skills are arranged hierarchically, wherein higher cognitive skill levels necessarily encompass lower cognitive skill lev-

els. Research validating the hierarchical nature of Bloom's taxonomy identifies one exception: the ordering of synthesis and evaluation (Kreitzer and Madaus, 1994; Anderson and Krathwohl, 2001).

Guiding Framework

Our efforts to characterize assessments are guided by prior research in the biological sciences that uses Bloom's taxonomy to objectively describe 1) assessments in pre- and postreformed courses (Freeman *et al.*, 2011; Haak *et al.*, 2011) and 2) the national landscape of assessments routinely used in introductory biology courses (Zheng *et al.*, 2008; Momsen *et al.*, 2010). In the undergraduate science classroom, the complex process of and motivations for learning are, it seems, inextricably tied to both instructional practices and the nature of assessments. We argue that if assessment truly drives learning, then it follows that the nature of our assessments can reveal a great deal about the cognitive skills our students are working to master.

METHODS

Course Descriptions

We conducted this study at a public university with very high research activity and an enrollment of nearly 12,000 students at the undergraduate level. Data were gathered from a single academic year (2010–2011) from two introductory science sequences, namely, introductory biology and introductory calculus-based physics.

Introductory biology is a large-enrollment, two-semester course sequence (referred to hereafter as Bio1 and Bio2) serving multiple science and engineering majors, in addition to prepping biology majors for upper-division coursework. Frequently, Bio1 serves as the first science class an undergraduate takes in college. Neither Bio1 nor Bio2 has pre- or corequisites (e.g., Bio1 is not a prerequisite for Bio2); however, Bio1 serve as a prerequisite for all upper-division coursework in biology, and Bio2 serves as a prerequisite for selected classes in organismal biology and ecology. Content addressed through Bio1 and Bio2 includes genetics, molecular biology, the chemistry of life, evolution, biodiversity, ecology, and plant and animal form and function. Classes meet weekly in either three 50-min lectures or two 75-min lectures. Instruction is best described as traditional: instructor-led lectures with some interactivity, through either clickers or think–pair–share (e.g., Prather and Brissenden, 2008). Assessment is primarily through multiple-choice midterm and final exams. We collected exams from a total of five sections of Bio1 and Bio2 (Table 3). Sections of Bio1 and Bio2 were team-taught, with a single instructor teaching for a specified period of time before passing the course to another instructor.

The introductory calculus-based physics sequence for science and engineering majors consists of two one-semester courses (referred to hereafter as Phys1 and Phys2): Classical Mechanics and Thermodynamics (Phys1) and Electromagnetism, Waves, and Optics (Phys2). Both courses are offered every Fall and Spring semester. Typically, approximately 80 and 180 students are enrolled each semester in Phys1 and Phys2, respectively. A few engineering major programs require an alternative course to Phys1 offered by the Mechanical

Table 2. Bloom's levels of cognitive skills, related competencies, and example questions^a

Cognitive level ^b	Bloom's level	Examples of competencies	Example questions, tasks ^c	
			Biology	Physics
LOC	Knowledge	Observation and recall of information	Seed germination is inhibited by _____ and promoted by _____. A. Ethylene, cytokinin B. Absciscic acid, gibberellins C. Cytokinin, ethylene D. Gibberellins, absciscic acid	In which process does the internal energy of an ideal gas NOT change? A. Isobaric compression B. Isochoric heating C. Isothermal expansion D. Adiabatic expansion E. None of the above.
	Comprehension	Understanding information, translating knowledge into new context, comparing and contrasting facts, predicting consequences	Eukaryotic genes are much larger than their corresponding mature mRNA-processed transcripts. This is because A. There are fewer mRNA nucleotides than DNA nucleotides. B. mRNA is single-stranded. C. DNA contains noncoding sequences called introns that are not part of the final mRNA product. D. mRNA containing noncoding sequences called introns that are not part of the final mRNA product.	A periodic plane wave is incident on a boundary between two different media. Suppose medium 1 is modified such that the propagation speed of the wave in that medium is increased by a factor of two. No other changes are made to the setup (e.g., the source of the wave remains unchanged). After the modification, does the wavelength of the transmitted wave <i>increase</i> , <i>decrease</i> , or <i>remain the same</i> ?
HOC	Application	Using information, especially in new situations, solving problems using skills or knowledge	Assuming independent assortment for all gene pairs, what is the probability that the following parent AaBbCc will produce a gamete of ABC? A. 1 out of 8 chances B. 1 out of 16 chances C. 3 out of 4 chances D. 9 out of 16 chances	A point charge Q_1 is at the origin and a point charge $Q_2 = +4\mu\text{C}$ is at $x = 4\text{ m}$. The electric field is zero at $x = 1\text{ m}$. Determine Q_1 .
	Analysis	Seeing patterns, recognizing hidden meanings, identifying and organizing system components	In every case, caterpillars that feed on oak flowers looked like oak flowers. In every case, caterpillars that were raised on oak leaves looked like twigs. These results support which of the following hypotheses? A. Differences in diet trigger the development of different types of caterpillars. B. Differences in air pressure, due to elevation, trigger the development of different types of caterpillars. C. The differences are genetic—a female will produce all flower-like caterpillars or all twig-like caterpillars. D. The longer day lengths of summer trigger the development of twig-like caterpillars.	Two pulses are incident from the left toward a free end of the spring, as shown. As the pulses reflect from the free end, could the magnitude of the maximum transverse displacement of the spring be twice the amplitude of each of the pulses? If so, during how many different time intervals will this occur? Explain.
	Synthesis	Using old ideas to create new ones, generalizing from given facts, relating knowledge from several areas	Draw a molecule that could be made using these elements and label the bonds as either polar covalent, nonpolar covalent, or ionic. If needed, an element can be used more than once to make your molecule.	

(Continued)

Table 2. Continued

Cognitive level ^b	Bloom's level	Examples of competencies	Example questions, tasks ^c	
			Biology	Physics
	Evaluation	Comparing and discriminating between ideas, assessing value of theories and hypotheses, making choices based on reasoned arguments, verifying value of evidence, recognizing subjectivity		

^aThese examples are by no means inclusive of all possible question types, but are representative of the questions comprising the data set used in this study.

^bLower-order cognitive level (LOC), higher-order cognitive level (HOC).

^cItems at all cognitive levels were not available from the sampled courses, as reflected by blank cells in the table.

Engineering Department, but all require Phys2, hence the difference in enrollment. The classes meet for four 50-min lectures every week for 16 wk in a large lecture hall. Weekly homework, typically a set of traditional end-of-the-chapter problems, is graded by a computerized system, LON-CAPA, on the basis of whether or not the final answer is correct. Students are not required to explain their reasoning or provide a written solution. Much like the Bio1/2 sequence, Phys1 and Phys2 do not have small-group discussion sessions. Lecture with some interactivity was the primary pedagogy used in class.

We collected exams from a total of four sections of Phys1 and Phys2 (Table 4). Depending on the instructor, exams were composed of either multiple-choice items exclusively or a combination of multiple-choice and free-response items (typically a 70/30 split). Free-response items required students to explain their reasoning and/or show their work. Each course was taught by a single instructor.

Rubric Development and Modification

Coding of assessment items from Bio1 and Bio2 used the rubric used by Momsen *et al.* (2010). The application of Bloom's taxonomy to Physics1/2 was a novel approach. A sample of physics assessment items (~100) from algebra-based physics was collected to simultaneously train the researchers on reliable application of Bloom's taxonomy to physics content and to modify the taxonomy to include examples and language reflective of the discipline. These items

were independently rated in batches of 20–30 items by each researcher and then compared and discussed in a group. With each coding and discussion session, modifying language and examples were added to the rubric. This final, revised Bloom's rubric was then used to rate the exam items collected from the calculus-based physics course.

Bloom's taxonomy categorizes items into increasingly higher cognitive skill levels. In practice, these cognitive skill levels are distributed across an ordinal scale, in which 1 = knowledge, 2 = comprehension, 3 = application, 4 = analysis, 5 = synthesis, 6 = evaluation (e.g., Wyse and Wyse, personal communication; Crowe *et al.*, 2008; Zheng *et al.*, 2008; Momsen *et al.*, 2010; Freeman *et al.*, 2011). Research validating the hierarchical nature of Bloom's taxonomy is somewhat mixed, with clear support for the ordering of levels 1–4, while the ordering of levels 5 and 6 is less straightforward, in part because there are few questions written at this level (Kreitzer and Madaus, 1994; Anderson and Krathwohl, 2001).

Interrater Agreement

Given the context-dependent nature of coding assessment items, we utilized two groups of raters composed of subject-matter experts. To ensure consistency in rating between biology and physics groups, a researcher with an expertise in biochemistry and a deep knowledge of both biology and physics was a member of both groups. Both groups followed published methodologies (Zheng *et al.*, 2008; Momsen *et al.*, 2010; Freeman *et al.*, 2011): three raters independently assigned a

Table 3. Course demographics for introductory biology

Section	Semester	Enrollment	Instruction	Number of	
				Assessments	Items
Bio1, Section 1	Fall 2010	420	Team-taught	4	178
Bio1, Section 2	Fall 2010	102	Single instructor	5	227
Bio2, Section 1	Fall 2010	321	Team-taught	4	224
Bio1, Section 1	Spring 2011	420	Single instructor	5	241
Bio2, Section 1	Spring 2011	234	Single instructor	6	230

Table 4. Course demographics for introductory physics

Section	Semester	Enrollment	Instruction	Number of	
				Assessments	Items
Phys1	Fall 2010	83	Single instructor	4	85
Phys2	Fall 2010	188	Single instructor	4	62
Phys1	Fall 2011	67	Single instructor	4	88
Phys2	Spring 2011	210	Single instructor	4	90

Bloom’s level to each assessment item and then met to discuss each question until a consensus rating was reached. To evaluate consistency between raters, we calculated agreement for each group of raters. Biology raters reached an agreement on 83% of items, and physics raters group reached an agreement of 76%. After a group discussion, consensus was reached on all items.

Student Performance Data

We report student performance data for each assessment item as the percentage of correct responses for a given item. We excluded “nonsense questions” that did not assess student understanding of science (e.g., “Which of the following individuals wrote your textbook?”) and poorly stated questions that could not be easily interpreted by the raters. Given the historical nature of this study, it was not possible to collect a complete data set for every assessment; however, our data include student performance from 21 assessments in biology collected in five independent course sections (Table 3) and 16 assessments in physics collected in four independent course sections (Table 4). Both multiple-choice and free-response items were coded; however, multiple-choice items were used exclusively for the correlation analysis between

the Bloom’s cognitive level and the item difficulty, as performance data from the free-response questions was not always readily available.

Statistical Analysis

To test for differences in cognitive levels assessed by introductory biology and university physics, we used chi-square analysis, and to quantify the magnitude of the difference between groups, we calculated effect size using Cramer’s *V*, which is an appropriate measure when working with categorical data and chi-square analyses (Smithson, 2003). We used Spearman’s rank correlation to test the alignment of Bloom’s cognitive level with student performance.

RESULTS

Introductory biology course assessment items were rated primarily at lower-order cognitive levels, knowledge (54% of items) and comprehension (39% of items), while introductory physics course assessments were rated primarily at comprehension (57% of items) and application (35% of items); see Figure 1 and Table 5. In physics, no items were rated as either evaluation or synthesis; in biology, one item was rated as synthesis, and no items were rated as evaluation. The distribution of questions in biology was significantly different from that in physics ($\chi^2(4) = 313.70, n = 1435, p < 0.001$) and the effect size was moderate ($V = 0.47, 95\% \text{ CI } [0.42, 0.52]$).

There is no significant relationship between student performance and Bloom’s level in introductory biology ($\rho = -0.018, df = 900, p = 0.59$); in physics, however, there is a weak but significant relationship between student performance and Bloom’s level ($\rho = -0.169, df = 276, p = 0.005$); see Figure 2.

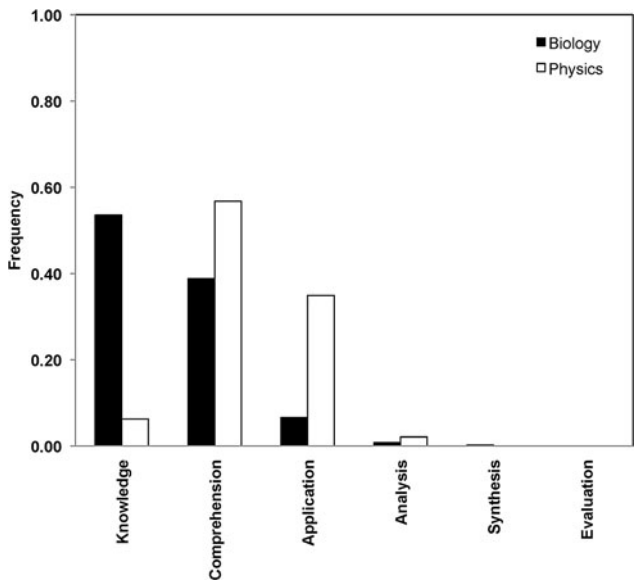


Figure 1. Distribution of assessment items by cognitive skill (Bloom’s) level for introductory biology and calculus-based introductory physics.

Table 5. Proportion of items categorized at each cognitive skill (Bloom’s) level

Bloom’s level	Biology	Physics
Knowledge	54%	6%
Comprehension	39%	57%
Application	7%	35%
Analysis	1%	2%
Synthesis	0.2%	0%
Evaluation	0%	0%

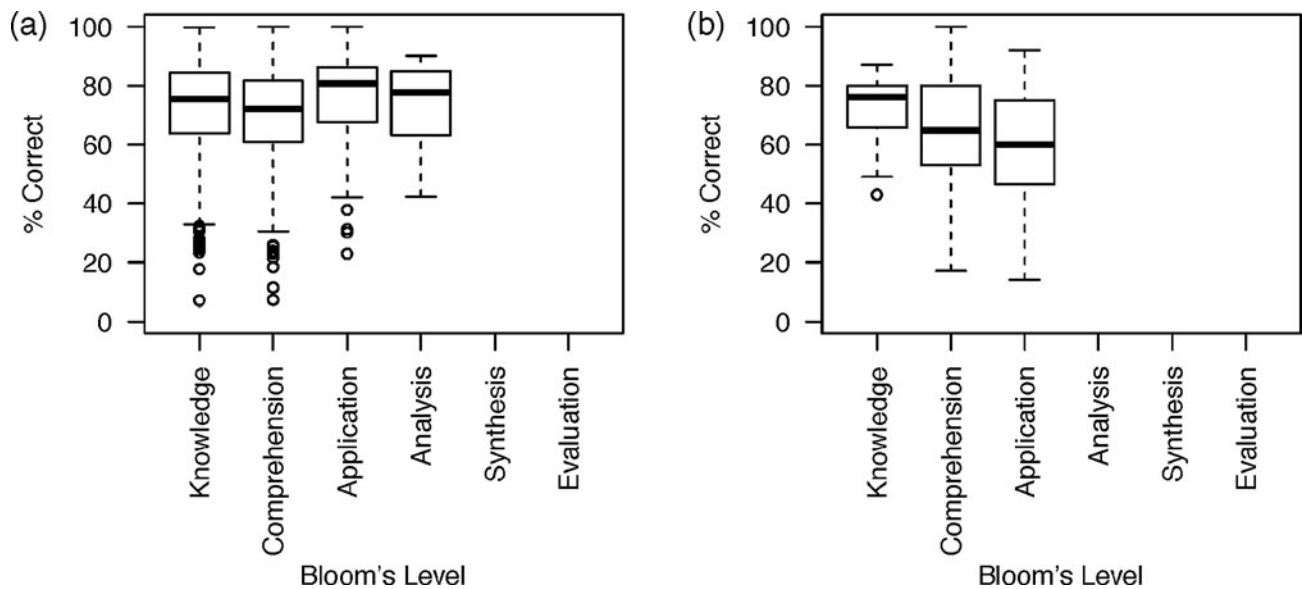


Figure 2. Relationship of cognitive skill (Bloom's) level to student performance for (a) introductory biology and (b) calculus-based introductory physics.

DISCUSSION

Our results suggest that, although physics and biology introductory course sequences assess lower cognitive skills, they do so at distinctly different levels. We argue that these differences may have profound implications for instruction related to 1) student expectations of and approaches to learning in a specific course and 2) assessing students' conceptual (rather than content) understanding. Furthermore, the introductory biology and calculus-based physics courses discussed here represent entry points for biology and physics students into their majors. As such, learning experiences in these courses may ultimately color students' perception of the major and impact their decisions to persist or leave.

Students' Epistemological Expectations

Students enter our courses with tacit beliefs about the nature of the knowledge to be learned, as well as the nature of successful learning strategies. These epistemological expectations are dynamic, influenced by contextual factors in the learning environment (Hall *et al.*, 2011). As one of many contextual factors, assessments have the potential to shape or reinforce students' epistemological expectations. Assessments that focus primarily on knowledge and comprehension are likely to reinforce expectations that understanding of a discipline is achieved by learning facts (Momsen *et al.*, 2010; Hall *et al.*, 2011). In response, students may adopt surface-learning approaches and avoid critically engaging with the content or practicing higher-order cognitive skills. For example, in biology, assessing predominantly knowledge and comprehension skills may discourage students from practicing and applying higher-order cognitive skills, such as synthesis and evaluation, in simplified biological systems. This is particularly troubling in introductory courses for biology majors, because students are likely to find themselves deficient in the skills needed to effectively solve problems in more complex

biological systems when they matriculate into upper-division coursework.

In contrast, one might argue that the examined introductory physics courses promote somewhat more sophisticated views on the nature of knowledge in science than the biology courses, because of the predominant emphasis of the physics assessments on higher levels of cognitive skills. Indeed, it is hoped that the application-level exam items, which focus on problem-solving skills, require students to 1) consider an unfamiliar situation previously not encountered in class; 2) articulate goals; 3) determine principles, concepts, and relations needed to analyze the presented situation; 4) translate conceptual understanding into numerical relationships; 5) apply specific skills relevant to a problem, such as drawing appropriate representations and transferring knowledge between representations; and 6) evaluate results (Maloney, 2011). Yet research on student epistemology suggests that, while some students do indeed engage in higher-order cognitive skills associated with sense-making and problem-solving processes, many students' epistemological expectations cause them to rely on previously successful approaches, such as searching for an equation to plug in numbers or memorizing and applying step-by-step problem-solving algorithms (Hammer, 1989; Hall *et al.*, 2011). For these students, this surface-level learning approach continues to be fruitful and reinforces epistemological expectations that a discipline consists of procedural knowledge and can be approached using algorithmic reasoning strategies. Therefore, the mere inclusion of application-level problems on exams does not guarantee student engagement in the desired deep-learning approach and may reinforce erroneous ideas about the nature of knowledge and learning within a discipline.

While there are differences between introductory biology and physics course sequences in terms of cognitive levels assessed, our data suggest that they are both limited to lower cognitive skills. As a result, the assessments routinely used in these courses may be reinforcing students' beliefs that

1) biology is about memorizing facts and physics is about plugging-and-chugging numbers into equations and 2) successful learning in both disciplines can be achieved through surface approaches, such as memorization and algorithmic reasoning strategies. As a result, we argue that Bloom's taxonomy is a potentially useful tool for instructors to more intentionally create assessments that challenge students' epistemological beliefs by assessing both lower and higher cognitive levels.

Cognitive Skill Level and Student Performance

The weak or missing relationship between cognitive skill level and student performance contradicts a common misconception in education at the K–12 and undergraduate level, namely, that cognitive complexity directly correlates with item difficulty (Lord and Bavisar, 2007; Wyse and Viger, 2011). Evidence from studies directly probing the relationship between cognitive complexity and difficulty underscores the limited or nonexistent relationship between these two constructs (Wyse and Wyse, personal communication; Nehm and Schonfeld, 2008). In fact, Bloom suggests that item difficulty is related to the content and context of a given problem (Bloom, 1956; Anderson *et al.*, 2002). For example, Nehm and Schonfeld (2008) found items that asked students to define a term or concept related to evolution (e.g., natural selection) were difficult (i.e., few students answered correctly) and that item context (e.g., evolution of plants) may have impacted item difficulty even further.

Therefore, while Bloom's taxonomy is a useful tool for categorizing the cognitive skills required to answer a question, it is not designed to measure other potential factors that may influence student performance, such as how students access cognitive resources or the metacognitive skills needed to recognize and apply appropriate reasoning strategies (Zimmerman *et al.*, 1992; Hammer *et al.*, 2005). Students may perform poorly on an assessment item not because of erroneous understanding, but rather because of how the problem was framed. Framing an item in a specific way may cause students to access inappropriate knowledge or reasoning strategies (Hammer *et al.*, 2005; Nehm and Ha, 2011). For example, in our study, the bulk of physics items assessed students at the comprehension and application levels, yet there was variability in student performance; some application items were easy for the majority of students, while others were more difficult. One interpretation of this observation is that students struggle to identify cognitive processes (e.g., reasoning strategies or an application of a specific concept) that are appropriate for the task at hand. Students lacking sufficient metacognitive skills might easily misapply a surface or algorithmic approach when higher-order skills were required, and vice versa. In biology, Nehm and Ha (2011) demonstrated a significant impact of item feature on students' evolutionary reasoning; for example, students were more likely to use core principles of natural selection when explaining trait gain and less likely to do so when explaining trait loss.

Within the domain of physics, conceptual questions were often rated at the comprehension level, and as a result might be erroneously thought of as easier by those equating cognitive level with item difficulty. However, research has demonstrated conceptual questions are notoriously difficult for a variety of reasons. In physics, many students tend to em-

ploy intuitive rather than formal reasoning, take shortcuts in their reasoning steps, and overly generalize results of simple cases to broadly applicable rules (Sabella and Cochran, 2004; Lising and Elby, 2005; Kryjevskaja *et al.*, 2011). In addition, one critical aspect of learning physics is making connections between a conceptual understanding and a mathematical description of a specific phenomenon. However, many students do not focus on analyzing how quantities that appear in math expressions are related to specific aspects of experimental setups. Instead, they tend to manipulate symbols, which often leads to erroneous conclusions. (Sherin, 2001; Loverude *et al.*, 2002; Kryjevskaja *et al.*, 2012). Similarly, conceptual questions in biology frequently require students to reason across levels of biological organization, yet students adopt a procedural rather than a conceptual reasoning approach (Bloome *et al.*, 1989; Jiménez-Aleixandre *et al.*, 2000). Abstract or implicit thought processes, such as the interplay between conceptual and mathematical reasoning skills characteristic for learning physics and connecting molecular mechanisms with emergent properties of a biological system, are not reflected in the Bloom's taxonomy of cognitive levels, yet they contribute to item difficulty.

Our findings relating item difficulty and cognitive level signal the need for further, more systematic research to fully characterize the diversity of factors affecting our abilities to accurately measure student learning. Bloom's taxonomy can reliably be applied to provide insights into the cognitive level of the activities we ask of our students, but is clearly limited in documenting the ways in which students are interpreting the context of assessment items, accessing cognitive resources, or applying reasoning strategies. Indeed, a recent study by Mesic and Muratovic (2011) investigating predictors of physics item difficulty suggests several factors that may influence item difficulty, including 1) complexity and automaticity of knowledge structures (the latter is one of the defining features of declarative knowledge, according to de Jong and Ferguson-Hessler [1996]), 2) the predominantly used type of knowledge representation, 3) nature of interference effects of relevant formal physics knowledge structures and corresponding intuitive physics knowledge structures (including p-prims), 4) width of the cognitive area that has to be "scanned" with the purpose of finding the correct solution and creativity, and 5) knowledge of scientific methods (especially experimental method).

Limitations

We recognize that exams are not the sum total of assessments in an introductory science course—homework assignments, in-class work, projects, and lab reports are also a substantial part of assessment and present significant learning opportunities for students. However, exams and quizzes are traditionally high stakes, representing a major portion of a student's final course grade and, as a result, significantly impact the learning approaches students adopt.

Further, this research represents a snapshot of assessments from multiple courses, instructors, and semesters at a single institution. Although our results align with other studies (Wyse and Wyse, personal communication; Zheng *et al.*, 2008; Momsen *et al.*, 2010), we recognize that institutions, departments, and instructors are diverse. As such, these results may not be representative of every introductory biology or physics

course. Indeed, as momentum to transform undergraduate science education builds (National Research Council, 1999, 2003; American Association for the Advancement of Science, 2011), we sincerely hope these results become baseline data, supporting research that documents the degree to which undergraduate education has evolved and improved.

Finally, our data were clustered around lower cognitive skill levels, with an emphasis on knowledge and comprehension in biology and comprehension and application in physics. Such clumping of data creates the potential for a floor effect. Although the data reflect in situ assessment practices, there is a need for more systematic investigations that look across all cognitive skill levels to further elucidate relationships between cognitive skill level and item difficulty and identify the multiplicity of factors affecting item difficulty, including student epistemologies. Such efforts may lead to a predictive model that can inform the development of assessment items to more accurately measure desired learning outcomes.

Implications for Teaching

Undergraduates' expectations of what it means to learn and engage in the process of science—their epistemologies—may impact how students learn by influencing course participation and interactions with course materials, including assessments (Hofer and Pintrich, 1997; Hofer, 2000, 2006; Limón, 2006). Students seem to expect that biology is a discipline focused on facts, devoid of deep understanding, and dissociated from a broader context and that introductory biology does not require knowledge of math and physics (Hall *et al.*, 2011). Physics students have similar epistemological ideas, namely, that physics is about facts and formulas and knowledge comes in largely disconnected pieces (Hammer, 1994a,b; Elby, 2011). Both physics and biology students perceive learning as absorbing knowledge from an authority (i.e., instructor or textbook) (Hammer, 1994a,b; Elby, 2011; Hall *et al.*, 2011). Indeed, it seems likely that assessments used in introductory biology and physics courses reinforce these expectations, through a focus on items that primarily test students' knowledge and comprehension, as in the case of biology, or comprehension and application, as in physics. As efforts to reform and transform undergraduate science courses mount, assessments send a clear message to students regarding the learning expectations for a given course and, as such, may play a critical role in helping students rebuild epistemologies that better align with disciplinary practices.

Our results provide science faculty with some insight into the expectations tacitly communicated to our students. Indeed, as our data demonstrate, there is a real and significant difference in the levels of cognitive skills demanded from students in two different science domains. We argue that, if students are likely to succeed in a biology course by memorizing and articulating ideas explicitly discussed in class, they have no basis for questioning learning strategies that have already proven effective. By extension, students entering a physics classroom are also likely to adopt similar surface-level approaches to learning physics. Increased instructor awareness of these differences may serve as an incentive to modify instruction in order to explicitly address this mismatch between students' and instructors' views on the development of expertise. Indeed, Redish and Hammer (2009) suggested

that implementation of reformed instruction that focuses on increased epistemological emphasis may be effective in addressing this mismatch. They found that explicit epistemological discussions incorporated into instruction, adaptation of peer instruction materials, and the use of epistemologically modified interactive lecture demonstrations not only yielded strong gains on conceptual tests, but also produced unprecedented gains on epistemological surveys rather than the traditional losses. Addressing the mismatch between students' epistemological beliefs about the nature of knowledge and learning in biology and physics (e.g., biology is about memorizing facts, and physics is about memorizing equations and plugging in numbers) and actual scientific practice is not commonly an explicit learning outcome of introductory science courses. As a result, students who struggle may be unable to identify, understand, and act on the underlying cause of their struggles, namely, the adoption of surface or strategic approaches, such as memorization or algorithmic problem solving, rather than engagement in sense making and logical development of ideas. As a consequence of this disparity, students exhibit poor performance, elevated levels of frustration, and increased attrition from the major.

Student learning is complex and influenced by factors perhaps too numerous to count. Exams and quizzes are used as the predominant mode of assessing student learning in introductory undergraduate science. The weak relationship between cognitive level and item difficulty highlights the need to carefully consider the ways in which we assess student learning. Due to the number of factors that are likely to impact the difficulty of an assessment item, we must be cautious in the conclusions we draw about student understanding and abilities when interpreting student performance data from a single source, such as exams. Our data lend support to national calls to integrate multiple and diverse measures of student learning to more accurately characterize the degree to which instruction is successful at achieving course outcomes (Huba and Freed, 2000). Indeed, understanding the cognitive skills routinely assessed in introductory biology and physics courses informs current assessment practices, while identifying potential pathways to modify and improve the assessment of undergraduates in science. Assessing at predominantly lower cognitive levels limits the development of critical reasoning and problem-solving skills that represent the heart and soul of scientific inquiry. If we truly wish to develop these skills, we must work to scaffold not just our instruction but also our assessments to include higher-order cognitive tasks.

ACKNOWLEDGMENTS

This research received approval from the local institutional review board under the protocols SM12217 and SM11247 and is supported in part by National Science Foundation Department of Undergraduate Education grant 0833268 to L.M.

REFERENCES

American Association for the Advancement of Science (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action*, Washington, DC.

- 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 (eds.) (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, New York: Addison Wesley Longman.
- Biggs JB (1987). *Student Approaches to Learning and Studying*, Hawthorn: Australian Council for Educational Research.
- Bloom BS (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook I: Cognitive Domain*, New York: Longmans, Green.
- Bloome D, Puro P, Theodorou E (1989). Procedural display and classroom lessons. *Curric Inquiry* 19, 265–291.
- Carter K (1984). Do teachers understand principles for writing tests? *J Teach Educ* 35, 57–60.
- Crooks TJ (1988). The impact of classroom evaluation practices on students. *Rev Educ Res* 58, 438–481.
- 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.
- de Jong T, Ferguson-Hessler M (1996). Types and qualities of knowledge. *Educ Psychol* 31, 105–113.
- Doyle W (1986). Classroom organization and management. In: *Handbook of Research on Teaching*, 3rd ed., ed. MC Wittrock, New York: Macmillan, 392–431.
- Elby A (2011). Getting started with research on epistemological expectations. In: *Getting Started in PER*, ed. C Henderson and KA Harper, College Park, MD: American Association of Physics Teachers.
- Entwistle A, Entwistle N (1992). Experiences of understanding in revising for degree examinations. *Learn Instruct* 2, 1–22.
- Entwistle N, Tait H, McCune V (2000). Patterns of response to an approaches to studying inventory across contrasting groups and contexts. *Eur J Psychol Educ* 15, 33–48.
- Entwistle NJ (1991). Approaches to learning and perceptions of the learning environment. *High Educ* 22, 201–204.
- Frederiksen N (1984). The real test bias: influences of testing on teaching and learning. *Am Psych* 39, 193–202.
- Freeman S, Haak D, Wenderoth AP (2011). Increased course structure improves performance in introductory biology. *CBE Life Sci Educ* 10, 175–186.
- Gijbels D, Segers M, Struyf E (2008). Constructivist learning environments and the (im)possibility to change students' perceptions of assessment demands and approaches to learning. *Instruct Sci* 36, 431–443.
- Guo SS (2008). Science education: should facts come first? *Science* 320, 1012–1012.
- Haak D, HilleRisLambers J, Pitre E, Freeman S (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science* 332, 1213–1216.
- Hall KL, Watkins JE, Coffey JE, Cooke TJ, Redish EF (2011). Examining the impact of student expectations on undergraduate biology education reform. *arXiv*, 1105.6349.
- Hammer D (1989). Two approaches to learning physics. *Phys Teach* 27, 664.
- Hammer D (1994a). Epistemological beliefs in introductory physics. *Cogn Instr* 12, 151–183.
- Hammer D (1994b). Students' beliefs about conceptual knowledge in introductory physics. *Int J Sci Educ* 16, 385–403.
- Hammer D, Elby A, Scherr RE, Redish EF (2005). Resources, framing, and transfer. In: *Transfer of Learning from a Modern Multidisciplinary Perspective*, ed. JP Mestre, Greenwich, CT: Information Age, 89–120.
- Hofer BK (2000). Dimensionality and disciplinary differences in personal epistemology. *Cont Educ Psych* 25, 378–405.
- Hofer BK (2006). Domain specificity of personal epistemology: resolved questions, persistent issues, new models. *Int J Educ Res* 45, 85–95.
- Hofer BK, Pintrich PR (1997). The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning. *Rev Educ Res* 67, 88–140.
- Huba ME, Freed JE (2000). *Learner-Centered Assessment on College Campuses: Shifting the Focus from Teaching to Learning*, Boston, MA: Allyn & Bacon.
- Jiménez-Aleixandre MP, Bugallo Rodríguez A, Duschl RA (2000). "Doing the lesson" or "doing science": argument in high school genetics. *Sci Educ* 84, 757–792.
- Krathwohl DR (2002). A revision of Bloom's Taxonomy: an overview. *Theor Pract* 41, 212.
- Kreitzer AE, Madaus GF (1994). Empirical investigations of the hierarchical structure of the taxonomy. In: *Bloom's Taxonomy: A Forty-Year Retrospective*, ed. LW Anderson and LA Sosniak, Chicago: University of Chicago Press, 64–81.
- Kryjevskaja M, Stetzer MR, Heron PRL (2011). Student understanding of wave behavior at a boundary: the limiting case of reflection at fixed and free ends. *Am J Phys* 79, 508–516.
- Kryjevskaja M, Stetzer MR, Heron PRL (2012). Student understanding of wave behavior at a boundary: the relationships among wavelength, propagation speed, and frequency. *Am J Phys* 80, 339–347.
- Limón M (2006). The domain generality–specificity of epistemological beliefs: a theoretical problem, a methodological problem or both? *Int J Educ Res* 45, 7–27.
- Lising L, Elby A (2005). The impact of epistemology on learning: a case study from introductory physics. *Am J Phys* 73, 372.
- Locke EA, Latham GP (1990). *A Theory of Goal Setting and Task Performance*, Englewood Cliffs, NJ: Prentice Hall.
- Lord T, Baviskar S (2007). Moving students from information recitation to information understanding: exploiting Bloom's taxonomy in creating science questions. *J Coll Sci Teach* 36, 40.
- Lortie DC (1975). *Schoolteacher: A Sociological Study of Teaching*, Chicago, IL: University of Chicago Press.
- Loverude ME, Kautz CH, Heron PRL (2002). Student understanding of the first law of thermodynamics: relating work to the adiabatic compression of an ideal gas. *Am J Phys* 70, 137.
- Maloney D (2011). An overview of physics education research on problem solving. Getting started in PER. *Rev PER* 2. www.compadre.org/Repository/document/ServeFile.cfm?ID=11457&DocID=2427.
- Marton F, Säljö R (1976). On qualitative differences in learning. II. Outcome as a function of the learner's conception of the task. *Brit J Educ Psych* 46, 115–127.
- Mesic V, Muratovic H (2011). Identifying predictors of physics item difficulty: a linear regression approach. *Phys Rev ST Phys Educ Res* 7, 010110.
- Micari M, Light G (2009). Reliance to independence: approaches to learning in peer led undergraduate science, technology, engineering, and mathematics workshops. *Int J Sci Educ* 31, 1713–1741.
- Momsen JL, Long TM, Wyse SA, Ebert-May D (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE Life Sci Educ* 9, 435–440.
- National Research Council (NRC) (1999). *Transforming Undergraduate Education in Science, Mathematics, Engineering, and Technology*, Washington, DC: National Academies Press.

- NRC (2003). *Bio 2010: Transforming Undergraduate Education for Future Research Biologists*. Committee on Undergraduate Biology Education to Prepare Research Scientists for the 21st Century, Washington, DC: National Academies Press.
- Nehm RH, Ha M (2011). Item feature effects in evolution assessment. *J Res Sci Teach* 48, 237–256.
- Nehm RH, Schonfeld IS (2008). Measuring knowledge of natural selection: a comparison of the CINS, an open response instrument, and an oral interview. *J Res Sci Teach* 45, 1131–1160.
- Prather EE, Brissenden G (2008). Development and application of a situated apprenticeship approach to professional development of astronomy instructors. *Astron Educ Rev* 7, 1.
- Prosser M, Trigwell K (1999). *Understanding Learning and Teaching: The Experience in Higher Education*, Philadelphia, PA: Open University Press.
- Redish EF, Hammer D (2009). Reinventing college physics for biologists: explicating an epistemological curriculum. *Am J Phys* 77, 629–642.
- Sabella MS, Cochran GL (2004). Evidence of intuitive and formal knowledge in student responses: examples from the context of dynamics. *AIP Conf Proc* 720, 89–92.
- Schunk DH (1990). Goal setting and self-efficacy during self-regulated learning. *Educ Psychol* 25, 71–86.
- Scouller K (1998). The influence of assessment method on students' learning approaches: multiple choice questions examination versus assignment essay. *High Educ* 35, 453–472.
- Scouller KM, Prosser M (1994). Students' experiences in studying for multiple choice question examinations. *Stud High Educ* 19, 267–279.
- Sherin BL (2001). How students understand physics equations. *Cogn Instr* 19, 479–541.
- Smithson M (2003). *Confidence Intervals, Quantitative Applications in the Social Sciences Series*, No. 140, Thousand Oaks, CA: Sage.
- Snyder B (1973). *The Hidden Curriculum*, Cambridge, MA: MIT Press.
- Stiggins RJ (2001). The unfulfilled promise of classroom assessment. *Educ Meas* 20, 5–15.
- Trigwell K, Prosser M (1991). Improving the quality of student learning: the influence of learning context and student approaches to learning on learning outcomes. *High Educ* 22, 251–266.
- Webb NL (1997). *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education*. Res MG No. 6, Washington, DC: Council of Chief State School Officers.
- Wyse AE, Viger SG (2011). How item writers understand depth of knowledge. *Educ Assess* 16, 185–206.
- Zheng AY, Lawhorn JK, Lumley T, Freeman S (2008). Assessment. Application of Bloom's taxonomy debunks the "MCAT myth." *Science* 319, 414–415.
- Zimmerman BJ, Bandura A, Martinez-Pons M (1992). Self-motivation for academic attainment: the role of self-efficacy beliefs and personal goal setting. *Am Educ Res J* 29, 663–676.