

Article

The Colorado Learning Attitudes about Science Survey (CLASS) for Use in Biology

Katharine Semsar,^{*†} Jennifer K. Knight,^{†‡} Gülnur Birol,^{§||} and Michelle K. Smith^{+‡¶}

^{*}Department of Integrative Physiology, University of Colorado, Boulder, CO 80309; [†]Science Education Initiative, University of Colorado, Boulder, CO 80309; [‡]Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309; [§]Biology Program, University of British Columbia, Vancouver, British Columbia, V6T 1Z3, Canada; and ^{||}Science Centre for Learning and Teaching, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada

Submitted October 25, 2010; Revised April 21, 2011; Accepted May 17, 2011
Monitoring Editor: Diane K. O'Dowd

This paper describes a newly adapted instrument for measuring novice-to-expert-like perceptions about biology: the Colorado Learning Attitudes about Science Survey for Biology (CLASS-Bio). Consisting of 31 Likert-scale statements, CLASS-Bio probes a range of perceptions that vary between experts and novices, including enjoyment of the discipline, propensity to make connections to the real world, recognition of conceptual connections underlying knowledge, and problem-solving strategies. CLASS-Bio has been tested for response validity with both undergraduate students and experts (biology PhDs), allowing student responses to be directly compared with a consensus expert response. Use of CLASS-Bio to date suggests that introductory biology courses have the same challenges as introductory physics and chemistry courses: namely, students shift toward more novice-like perceptions following instruction. However, students in upper-division biology courses do not show the same novice-like shifts. CLASS-Bio can also be paired with other assessments to: 1) examine how student perceptions impact learning and conceptual understanding of biology, and 2) assess and evaluate how pedagogical techniques help students develop both expertise in problem solving and an expert-like appreciation of the nature of biology.

INTRODUCTION

Over the last decade, there has been a strong call to improve biology undergraduate education (e.g., Howard Hughes Medical Institute, 2002; National Research Council, 2003; Handelsman *et al.*, 2004; Wood, 2009; Woodin *et al.*, 2009). The main goals of this charge are to advance students' conceptual content knowledge to a deeper, more expert-like level

and to help students develop both expert-like approaches to problem solving and more sophisticated perceptions about how biology knowledge is structured. An important step in achieving these goals is to create assessments that measure whether curricular and pedagogical changes in the classroom are succeeding in both improving student learning and transitioning students toward more expert-like thinking (reviewed in Garvin-Doxas *et al.*, 2007; Woodin *et al.*, 2009; Knight, 2010).

DOI: 10.1187/cbe.10-10-0133

[¶]Present address: School of Biology and Ecology, University of Maine, Orno, ME 04496

Address correspondence to: Jennifer K. Knight (knight@colorado.edu).

© 2011 K. Semsar *et al.* 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.

While many researchers have focused on creating new assessments for conceptual understanding for specific subdisciplines of biology (e.g., Anderson *et al.*, 2002; Klymkowsky *et al.*, 2003; Bowling *et al.*, 2008; Garvin-Doxas *et al.*, 2007; Smith *et al.*, 2008; Shi *et al.*, 2010), there is a separate need to measure students' approaches to problem solving and understanding of the nature of biology as a scientific discipline. These epistemological assessments are necessary because educational reform that can improve conceptual learning does not necessarily also improve students' trajectories toward more expert perceptions of the field (Redish *et al.*, 1998; Adams *et al.*, 2006; Perkins *et al.*, 2005; Moll and Milner-Bolotin, 2009). Moreover, preheld beliefs about a discipline

can strongly impact conceptual learning by affecting how a student approaches learning and makes sense of a particular area of study (House, 1994, 1995; Perkins *et al.*, 2005). Therefore, it is critical that we learn how our educational practices impact the underlying perceptions students have about biology.

One way to examine student perceptions about biology is on a continuum of novice-to-expert level. Hammer (1994) proposed that differences between how experts and novices view a discipline can be characterized into three main areas: 1) content and structure of knowledge, 2) source of knowledge, and 3) problem-solving approaches. In the first area, content and structure of knowledge, experts believe knowledge is structured around a coherent framework of concepts, while novices believe knowledge comprises isolated facts that are not interrelated. In the second area, source of knowledge, experts believe knowledge about the world is established by experiments that describe nature, while novices believe knowledge is handed down by authority and do not see a connection to the real world. Finally, with regard to problem-solving approaches, experts rely on concept-based strategies that are widely applicable to multiple problem-solving situations, while novices will often apply pattern-matching to memorized problems and focus on surface features, rather than underlying concepts. Examination of all three of these areas demonstrates that experts not only have a deeper conceptual knowledge of a discipline, but they also hold more sophisticated views about how scientific knowledge is obtained, expanded, and structured, as well as how to approach unsolved problems.

To specifically examine how students perceive the field of biology on this expert-to-novice continuum, the Colorado Learning Attitudes about Science for Physics (CLASS-Phys; Adams *et al.*, 2006) was modified for biology. The CLASS-Phys survey was designed to probe student perceptions encompassing the three types of novice/expert thinking mentioned above and to allow for a direct comparison of student perceptions to expert perceptions (Adams *et al.*, 2006). CLASS-Phys was also recently modified for chemistry (CLASS-Chem; Barbera *et al.*, 2008). Like CLASS-Phys and -Chem, CLASS-Bio is designed for use in a wide range of university courses in the field of biology (major and nonmajor courses; lower- and upper-division courses). Evidence of response validity was collected through interviews with students and experts (biology PhDs) from various subdisciplines within the field of biology. This paper describes CLASS-Bio's design, evidence of its instrument validity, and results of its initial use in classrooms.

INSTRUMENT DESIGN AND EVIDENCE OF VALIDITY

General CLASS Characteristics and Benefits

All CLASS instruments (for physics, chemistry, and now biology) have been designed to compare novice and expert perceptions about the content and structure of a specific discipline; the source of knowledge about that discipline, including connection of the discipline to the real world; and problem-solving approaches. Respondents are asked to rate their agreement to each survey statement (e.g., "I think of

biology in my everyday life.") on a five-point Likert scale ("strongly agree" to strongly disagree). All CLASS-Bio statements are listed in Supplemental Material A. This format typically allows respondents to complete the instrument in less than 10 min. Likewise, scoring by instructors is straightforward with the aid of Excel scoring templates.

CLASS instruments are also unique in their ability to specifically detect novice/expert differences. Specifically, CLASS instruments compare student responses with the consensus response of experts (PhDs in the field). Only statements with an unambiguous and nearly unanimous expert response are included on CLASS instruments (see Expert Consensus below). Furthermore, previous work demonstrated that students can correctly identify expert responses to CLASS instrument statements but will still answer honestly about their own perceptions, regardless of whether they agree or disagree with the expert response (Adams *et al.*, 2006; Gray *et al.*, 2008). These results provide support that the CLASS instruments accurately reflect student thinking and not what students perceive their instructors want them to think.

While CLASS statements were initially based upon the Maryland Expectations (MPEX) Survey and Views about Science Survey (VASS), they differ from those and other surveys, such as the Views of the Nature of Science (VNOS) Survey and Student Assessment of Learning Gains (SALG), because they are not course-specific and measure perceptions about the discipline rather than self-efficacy (Adams *et al.*, 2006). In addition, all statements have been carefully worded and reviewed through student interviews to ensure that each statement has a single interpretation that is meaningful to students regardless of whether they have taken courses in the discipline. These features make the CLASS instruments especially useful when trying to track student perceptions across a curriculum and across courses (Adams *et al.*, 2006; Gire *et al.*, 2009). Furthermore, CLASS is unique among epistemological assessments, because categories of student thinking (Real World Connections, Enjoyment [Personal Interest], etc.) are based on statistical analysis of student responses, rather than categories determined by the instrument designers. The strength of this approach is that the categories are based on student data, rather than on what researchers believe are logical groupings.

Administration and Scoring

The scoring and administration of CLASS-Bio follow the same protocols determined for the CLASS-Phys and -Chem surveys (Adams *et al.*, 2006; Barbera *et al.*, 2008). The instrument is given to students online at the beginning and end of the semester, and students receive participation points (specific credit determined by instructor) for completing the survey. Student response rate for CLASS-Bio is high: the average response rate from seven classes across two different institutions is 75% for preinstruction surveys and 70% for postinstruction surveys.

Because it is difficult to control how seriously students take online assessments, it is important to exclude responses where students are unlikely to be reading the questions. The criteria used for response rejection include taking less than 3 min to complete the survey, providing the same Likert-scale response (for example all "strongly agree") for more than 90% of the statements, and incorrectly responding to a statement

embedded in the survey (“We use this statement to discard the survey of people who are not reading the questions. Please select agree (not strongly agree) for this question to preserve your answers.”). In addition, if a student submits more than one set of acceptable responses, only the first response is scored. Finally, only responses from students who fulfill all survey requirements on both the pre- and postinstruction surveys are included for analysis.

To score student responses, each response was collapsed from a five-point Likert scale to a three-point scale: “neutral,” “agree” (“strongly agree” or “agree”), or “disagree” (“strongly disagree” or “disagree”). Combining the “strongly agree”/“agree” and “strongly disagree”/“disagree” responses is consistent with an ordinal characterization of a Likert scale that assumes differences between responses on the scale are not necessarily equal. Interviews from both the initial CLASS-Phys validation and our own interviews support this interpretation: student reasons for answering “agree” versus “strongly agree” or “disagree” versus “strongly disagree” were not consistent from student to student. For example, for Statement 19, “The subject of biology has little relation to what I experience in the real world,” students with different responses offered similar reasoning: “disagree, again biology is the world”; “strongly disagree, because biology is our real world”; “disagree, I interact with biology all the time”; “strongly disagree, it is everywhere.” (See also Adams *et al.* [2006] for a deeper discussion of this issue.) Each student’s response is then designated as favorable (agreeing with the expert consensus—not necessarily agreeing with statement), unfavorable (disagreeing with expert consensus), or neutral (neither agreeing or disagreeing with the expert consensus). For example, for Statements 1 and 2 (Supplemental Material A), the expert response is “agree” and only students scoring “agree”/“strongly agree” are scored as having a favorable response. Meanwhile, for Statements 3 and 4 (Supplemental Material A), the expert consensus is “disagree” and, thus, only students answering “disagree”/“strongly disagree” are scored as having a favorable response.

Finally, an overall percent-favorable response score for each student is calculated that represents the percentage of statements on which the student had a response identical to the consensus response of experts. These scores are averaged to obtain an overall score for all respondents. Similarly, the scores can be subdivided by category, so that various aspects of student thinking can be examined separately (e.g., Real World Connections or Enjoyment [Personal Interest]).

Creating CLASS-Bio

CLASS-Bio was created following the same procedures used to adapt CLASS-Phys for Chemistry (Barbera *et al.*, 2008; Table 1). CLASS-Phys, -Chem, and -Bio statements were modified, added, or dropped based on faculty working-group discussions, student interviews, expert reviews, and factor analysis. Evidence of response validity was collected for individual statements through both student interviews and expert reviews. A pilot version of CLASS-Bio was given in Fall 2007, and after further modification, a final version in Fall 2008. After both versions were administered, CLASS-Bio statements were grouped into categories of student thinking based on results of a factor analysis on responses from large

Table 1. Sequence of CLASS-Bio statement development

1.	Examined CLASS-Phys and -Chem for statements that could apply to CLASS-Bio
2.	Met with faculty working groups to determine which statements should be included on CLASS-Bio
3.	Interviewed students and made modifications to statements based on their responses
4.	Solicited expert opinions and responses to statements
5.	Gave pilot version of CLASS-Bio (Fall 2007) and performed factor analysis following procedures listed in Table 2 (from Adams <i>et al.</i> , 2006)
6.	Revised statements and solicited more feedback from faculty working groups, student interviews, and experts
7.	Administered final version of CLASS-Bio (Fall 2008) and performed a second independent factor analysis (again following all procedures in Table 2)
8.	Verified category robustness using a polychoric correlation matrix-based factor analysis (see text for details)

introductory biology lecture courses. Finally, reliability and concurrent validity (whether the instrument can detect expected differences between populations) were measured for the final version of CLASS-Bio.

Faculty Working Group. To determine which statements should be included on CLASS-Bio, faculty working groups at University of Colorado at Boulder (CU) and University of British Columbia (UBC) first examined the CLASS-Phys and CLASS-Chem statements. Faculty selected statements ($n = 16$) where the word “biology” could be substituted for either “physics” or “chemistry.” For example, “Learning [biology] changes my ideas about how the natural world works.” They also modified CLASS-Phys and -Chem statements ($n = 13$) to reflect differences in the field of biology by de-emphasizing terms like “equations” or “solving problems” and replacing them with terms like “principles” or “ideas.” For example, the CLASS-Phys and -Chem statement “I do not expect equations to help my understanding of the ideas in chemistry; they are just for doing calculations” was changed to “I do not expect the rules of biological principles to help my understanding of the ideas.” Seven of the CLASS-Phys and -Chem statements were not included on CLASS-Bio because faculty found them inappropriate for the whole field of biology. Finally, the faculty working groups also generated new statements that were added to CLASS-Bio (CU, $n = 14$; UBC, $n = 7$). For example, “It is a valuable use of my time to study the fundamental experiments behind biological ideas.”

Student Interviews. To verify that wording and meaning of statements are clear to students and that there is consistency between students’ Likert-scale responses and verbal explanations, 39 students were interviewed (22 majors, 17 nonmajors/undeclared; 10 males, 29 females). At CU, 22 students were interviewed; 14 on the Fall 2007 pilot version, and an additional eight on the final Fall 2008 version. In addition, at UBC, seven statements were developed by faculty and then validated with 17 student interviews.

During each interview, students first completed a paper version of the survey and submitted it to the interviewer. Students were then verbally presented with each of the statements and asked to provide their response along with an explanation of the response. For each statement on the

final version of CLASS-Bio, an average of 85% of the verbal answers matched written responses. Among those that did not match, most were changes between “neutral” and either “agree” or “disagree” or vice versa. Only 2% of all responses changed from “agree” to “disagree” or vice versa.

Most students freely provided their thoughts and comments on all statements without any further prompting. If a student sat in silence for approximately 30 s without providing an explanation, he or she was prompted to explain his or her choice. Every interview was audiotaped and transcribed. In general, students were able to provide explanations consistent with their answer choices. However, for instances where multiple students identified ambiguity in a statement, the statement was revised or eliminated from the survey. For example, one statement “I can usually figure out a way to solve biology questions” resulted in ambiguity centered around whether students were thinking of doing questions on tests or on homework. Nearly one-half of students interviewed felt that they could easily look up answers outside of class. As “looking up answers” was not the intention of the statement, we dropped this statement from the survey. Explanations of how other statements were modified following student interviews are provided in Supplemental Material B.

Expert Consensus. To define the expert response and receive additional feedback on CLASS-Bio statements, biology experts were asked to take the survey online. An expert was defined as someone with a PhD in any field of biology. We obtained survey responses from 69 experts representing 30 universities: 41% molecular biologists, 26% physiologists, 20% ecologists, and 13% other subdisciplines. We asked experts to rate each statement on a Likert scale and gave them the opportunity to provide written comments on each statement. On the final Fall 2008 version of CLASS-Bio, we received an average of 49 expert ratings per statement (max = 69, min = 21). If fewer than 70% of the experts did not select either “strongly agree” / “agree” or “strongly disagree” / “disagree” to a particular statement, the statement was dropped from the instrument. If the faculty average consensus opinion was neutral for a statement, it was also dropped from the survey. The average consensus of opinion for all the statements on the final version of CLASS-Bio is 90%.

Factor Analysis and Statement Characterization. To examine the results of CLASS-Bio beyond the overall percent agreement with the experts’ score, we categorized the statements into different aspects of student thinking. The CLASS instruments are unique among epistemological assessments in that the statement categories are based on student responses. Other instruments have categories based on what researchers believe are reasonable groupings. In contrast, CLASS categories were determined by performing an iterative reduced-basis factor analysis on a large data set of student responses (where the number of student responses is 10 times greater than the number of statements) to find statistically valid categories. This process is described in detail in Adams *et al.* (2006) and outlined in Table 2.

As in the other CLASS instruments, the main goal of categorizing statements is to be able to examine student thinking in more depth, rather than to explain the variance of the overall scores. Therefore, we performed a factor analysis that allows statements to fall into multiple categories.

Table 2. Abbreviated methods of category development using iterative reduced-basis factor analysis (from Adams *et al.*, 2006)

1.	Data transformation: Data from a large data set was transformed to 3-point scale (Agree with expert, Neutral, Disagree with expert).
2.	Factor analysis: To find statistically valid categories of student thinking, a factor analysis was run in PASW (SPSS, Chicago, IL) with the following parameters: Correlation matrix: Pearson; Extraction Method: Principal Component Analysis; Rotation Method: Direct Oblimin; Cross-loading: Allowed.
3.	Category analysis and revision: For each potential component (hereto referred to as category), the effects of individual statements on the category were analyzed, such that if category statistics became stronger by either removing a statement or adding a potentially related statement, the category composition was changed to reflect the strongest statistics. The statistics used to make these judgments were the strength and similarity of factor loadings (contribution of each statement to that category), strength and similarity of statement correlations, and the linearity of the scree plot (an indication of whether a single component accounted for the variability in the statement group).
4.	Robustness indicators: An RI based on the aforementioned statistics was calculated for each potential category. Robustness indicators range from 0–10 and scores higher than 5 represent both high factor loading and high statement correlations, and thus a meaningful grouping of student thinking
	$\text{Robustness} = (2cc + fl + 5 \Delta E /N) \times 3R^2$
	where cc = average absolute value of the correlation coefficients between statements; fl = average absolute value of factor loadings; ΔE = shape of scree plot; N = number of statements in category; R^2 = Pearson product moment correlation (the linearity of a scree plot).
5.	Category names: Final categories were named according to expert opinion of the commonalities among the composition of statements within each category.

Final categories were chosen using a robustness indicator (RI; developed by Adams *et al.*, 2006), a measure of how strongly statements group together and, thus, how closely related student thinking is on the statements (see Table 2 for equation). RIs range from 0 to 10 (10 being most robust), and a score of at least 5 (our cutoff) indicates that both statement correlations and factor loadings (each statement’s contribution to the category) are high and statements are evenly contributing to a single coherent category. Following the final outcomes of the Fall 2008 analysis, each category was given a name representative of the nature of statements that fell within each statistically defined and response-based group (Table 3).

Compared with Adams *et al.* (2006), there were two major differences in our approach for CLASS-Bio. First, rather than performing an exploratory and then confirmatory factor analysis, we performed two independent analyses, the first on the Fall 2007 pilot version data and the second on the Fall 2008 version data. Second, we added a verification of each category’s robustness using data derived from a polychoric correlation matrix (designed for ordinal data). The original factor analyses were designed to be consistent with the previous methods used in the development of CLASS-Phys

Table 3. CLASS-Bio statement categorization and RIs^a

CLASS category	Statements	RI: Pearson	RI: Polychoric
Real World Connection	2, 12, 14, 16, 17, 19, 25	6.74	8.26
Enjoyment (Personal Interest)	1, 2, 9, 12, 18, 27	10.0	10.0
Problem-Solving: Reasoning	8, 14, 16, 17, 24	6.57	7.38
Problem-Solving: Synthesis & Application	3, 5, 6, 10, 11, 21, 30	7.10	8.96
Problem-Solving: Strategies	7, 8, 20, 22	7.14	7.09
Problem-Solving: Effort	8, 12, 20, 22, 24, 27, 30	6.62	7.53
Conceptual Connections/Memorization	6, 8, 11, 15, 19, 23, 31, 32	5.61	7.19
Uncategorized questions	4, 13, 26, 29	n/a	n/a

^a Statements in bold appear on CLASS-Phys and CLASS-Chem (in either the same or slightly modified forms), although not necessarily in the same categories. RIs, calculated with either the Pearson or polychoric correlation matrices, range from 0 to 10, with 10 being most robust.

and CLASS-Chem (Adams *et al.*, 2006; Barbera *et al.*, 2008) and included using a Pearson correlation matrix (designed for continuous data). However, as polychoric correlation matrices are more appropriate for ordinal data (Panter *et al.*, 1997), we ran a new factor analysis using a polychoric correlation matrix (SAS Software, Cary, NC), calculating new RIs for each of our previously determined categories. All of the categories identified using the Pearson correlation matrix factor analysis were verified (often with stronger RIs) using the polychoric correlation matrix–based factor analysis (Table 3).

The initial factor analysis on student responses was performed on responses from the Fall 2007 pilot version of CLASS-Bio, administered to students in two large Ecology and Evolutionary Biology (EBIO) introductory biology courses ($n = 627$). The second independent factor analysis was performed on responses from the final version of CLASS-Bio statements administered to students in the same courses in Fall 2008 ($n = 673$). Results from the Fall 2007 factor analysis resulted in 26 statements grouped into six robust and overlapping categories. Eighteen statements did not fall into any robust categories. While 14 of these were dropped from the instrument, four statements remained because they examined unique ideas and expert comments suggested that these statements were of general interest to faculty. The addition of student responses ($n = 214$) from a CU Molecular Cellular and Developmental Biology (MCDB) introductory biology course into the data set did not alter the results of the initial factor analysis (unpublished data), so only EBIO student data were used for the analyses.

The final version of CLASS-Bio was administered in Fall 2008 to the same EBIO course and included 30 of the initial statements and eight additional statements (one modified from a CLASS-Phys/-Chem statement and seven new statements developed by faculty from CU and UBC). The second independent factor analysis on Fall 2008 responses led to similar categories as determined for the Fall 2007 version and the generation of one additional robust category (Problem-Solving: Reasoning). Of the eight additional statements that were added to the survey in Fall 2008, only two grouped into categories and the other six were dropped. These final categories were verified using an additional polychoric correlation matrix–based factor analysis on the same Fall 2008 version data.

The final CLASS-Bio categories and their RIs are listed in Table 3. The final version of CLASS-Bio contains 31 state-

ments: 10 statements are identical to statements on CLASS-Phys and -Chem with the word “biology” substituted for “physics” or “chemistry,” 11 statements are modified from CLASS-Phys, and 10 statements are new and not found on the other CLASS instruments. (Supplemental Material A lists all of CLASS-Bio statements and whether they are identical or modified from the CLASS-Phys and -Chem surveys.) Final category RIs on CLASS-Bio (Pearson correlation matrix range = 5.61–10.0; polychoric correlation matrix range = 7.09–10.0; Table 3) are comparable to those on CLASS-Phys (range = 5.57–8.25; Adams *et al.*, 2006) and CLASS-Chem (range = 6.01–9.68; Barbera *et al.*, 2008).

While the modification of CLASS-Phys for Chemistry resulted in categories with highly similar statement groups (Barbera *et al.*, 2008), the categorization of CLASS-Bio statements differs from the previous CLASS surveys. For example, comparing CLASS-Chem categories to the original eight categories on CLASS-Phys, three categories have identical statement composition, three have just one or two novel CLASS-Chem statements added to an original CLASS-Phys statement composition, and only two categories dropped an original statement and added one or two new statements. In contrast, no CLASS-Bio category has more than 50% of the statements in any of the CLASS-Phys categories, even among categories that do not include any of the newly developed CLASS-Bio statements. As with any factor analysis, any change in the suite of statements can change the likelihood that statements will group together, and this is likely the main source of difference here. In the development of CLASS-Chem, 39 of 42 original CLASS-Phys statements were used (37 were kept the same [with “chemistry” replacing “physics”] and two were reworded) and 11 new statements were added. In contrast, on CLASS-Bio, only 21 of the 42 original CLASS-Phys statements were used (only 10 were kept the same [with “biology” substituting for the discipline name] and 11 were reworded [largely to allude to solving “problems” rather than “equations”]), and 10 new statements were added (for details see Supplemental Material A). Thus, the statement pool for CLASS-Bio was quite different from the other two CLASS instruments.

Evidence of Reliability and Validity for the Complete Instrument. Because of the large ($n > 600$) and consistent student population in the EBIO introductory biology courses (admissions standards and student populations

characteristics are nearly identical year-to-year), the reliability of the instrument was tested by calculating a test-retest coefficient of stability on student responses from two equivalent populations (Crocker and Algina, 1986; American Education Research Association, American Psychological Association, National Council on Measurement and Education, 1999). This measure is preferable to Cronbach's alpha in this case for two main reasons: 1) because it measures stability over time rather than internal consistency, and 2) Cronbach's alpha was designed for use with single-construct assessments (assessments that test one main concept), while CLASS-Bio addresses multiple constructs, such as structure of knowledge, source of knowledge and problem solving (Crocker and Algina, 1986; Cronbach and Shavelson, 2004; Adams and Wieman, 2011). The coefficient of stability was calculated by correlating all preinstruction responses between students in the Fall 2007 and Fall 2008 courses (representing 100% of the final 31 survey statements on CLASS-Bio). The coefficients of stability for CLASS-Bio are as follows: percent-favorable responses, $r = 0.97$; percent neutral responses, $r = 0.91$; percent unfavorable responses $r = 0.97$. As $r = 0.80$ is considered an indication of high reliability, these results support a high reliability of CLASS-Bio and are consistent with results from the reliability analyses of CLASS-Phys (respectively: $r = 0.99, 0.88, \text{ and } 0.98$; Adams *et al.*, 2006) and CLASS-Chem (respectively: $r = 0.99, 0.95, \text{ and } 0.99$; Barbera *et al.*, 2008).

Finally, similar to CLASS-Phys and -Chem (Adams *et al.*, 2006; Barbera *et al.*, 2008) and as further evidence of instrument validity, the concurrent validity of CLASS-Bio was examined by determining whether the instrument can detect differences between student populations that would be plausibly expected to hold different perceptions. The CLASS-Bio overall percent-favorable scores between declared biology majors and declared nonmajors (largely psychology majors) were compared within the same EBIO introductory biology course in Fall 2008. The declared biology majors in a large introductory biology course had significantly higher percent-favorable scores (which indicate percent agreement with experts) entering the course than nonmajors in all but one category (Figure 1). In addition, another study comparing majors and nonmajors in CU's MCDB genetics courses shows similar results with student in the majors' course having higher percent-favorable scores than students in the nonmajors course (Knight and Smith, 2010).

Approval to evaluate student responses and interview students was granted by the Institutional Review Board, CU (expedited status, Protocol No. 0603.08), and the Behavioural Research Ethics Board, UBC (Protocol No. H07-01633).

CLASSROOM STUDIES

We have given the final version of CLASS-Bio to students in both introductory and upper-division biology courses across four institutions (CU; UBC; Western Washington University, Bellingham, WA; and University of Maryland, College Park, MD). Data from CU's Department of Integrative Physiology (IPHY), MCDB, and EBIO, and UBC's biology program are presented here.

Figure 2A shows the overall pre- and postinstruction percent-favorable scores (percent agreement with experts) in

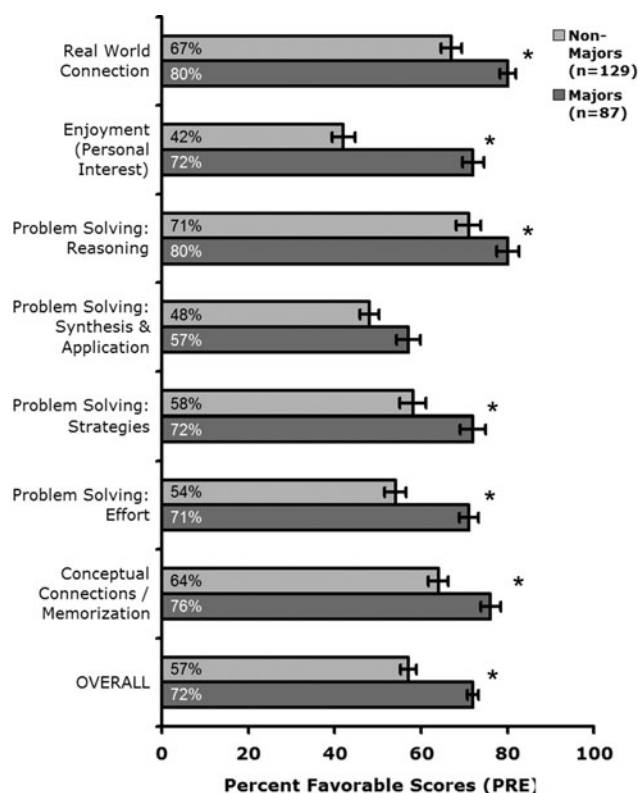


Figure 1. Differences in CLASS-Bio percent-favorable scores between majors and nonmajors entering an introductory biology course. Percent-favorable scores are measures of percent agreement with the experts (see text for details). Asterisks indicate that majors have significantly higher scores entering an introductory course than nonmajors in that category (>2 SEM).

introductory biology courses in two CU departments (EBIO and MCDB) and one UBC program (biology). (Courses are coded A–F for anonymity.) In five of these six introductory courses, there is a significant decrease in the student-matched percent-favorable score across the instruction period (course averages of the paired student data between pre- and postinstruction are considered significantly different if they differ by more than 2 SEM; Adams *et al.*, 2006). In other words, students become more novice-like in their beliefs during their introductory biology courses. Examples of individual category shifts from two introductory courses are displayed in Figure 3, A and C.

In contrast, student CLASS-Bio overall percent-favorable scores in five upper-division courses in two CU departments show no overall shifts across the instruction period (Figure 2B). This trend holds true when the statements are divided into categories (Figure 3, B and D). Furthermore, in one upper-division course, there were significant expert-like shifts in specific categories at the end of the instruction period (Figure 3B).

Finally, students entering upper-division courses have comparable or higher preinstruction scores than students' postinstruction scores in introductory courses (examples of two curriculum series are shown in Figure 3, A–B and C–D).

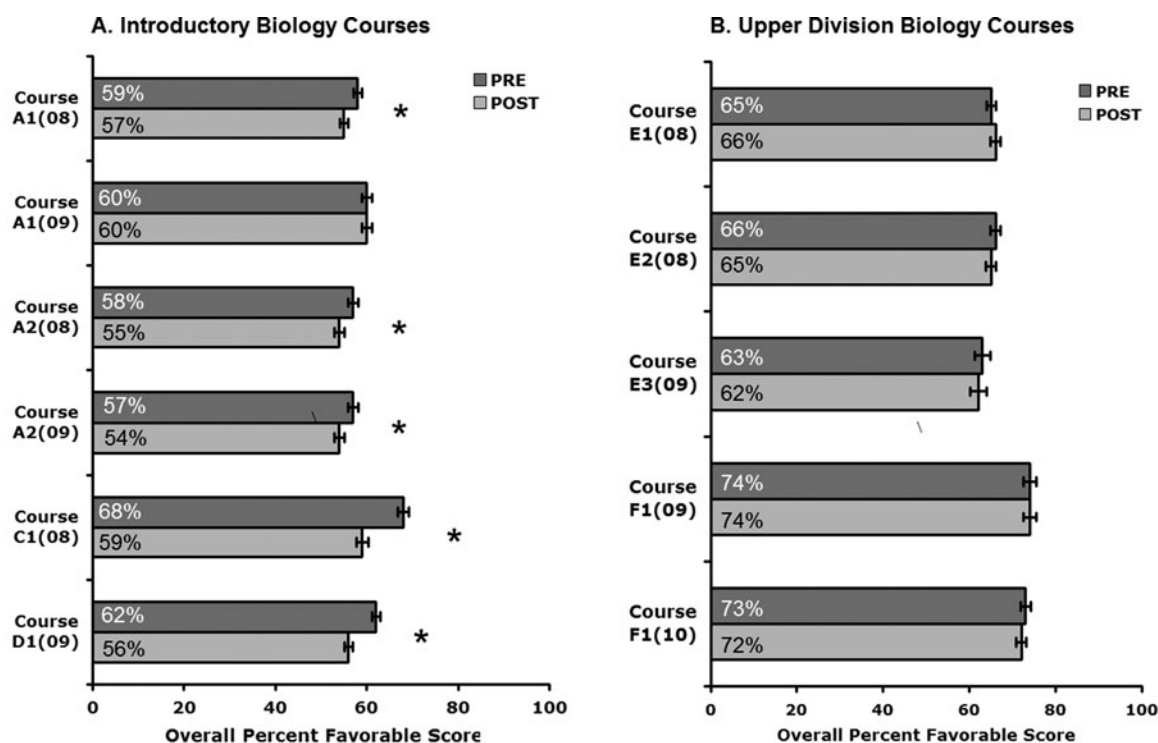


Figure 2. Overall pre- and postinstruction percent-favorable scores (percent agreeing with expert) in introductory (A) and upper-division (B) courses. Courses are coded by course (letter), instructor (number), and year. Introductory courses are represented by two CU departments (EBIO and MCDB) and one UBC department (biology) while upper-division courses are represented by two CU departments (MCDB and IPHY). Sample sizes are as follows: A1(08), $n = 370$; A1(09), $n = 336$; A2(08), $n = 287$; A2(09), $n = 265$; C, $n = 170$; D, $n = 504$; E1, $n = 126$; E2, $n = 130$; E3, $n = 126$; F(09), $n = 81$; F(10), $n = 79$. Asterisks indicate significant differences between pre- and postinstruction scores based on paired student data (>2 SEM).

DISCUSSION

CLASS-Bio Design and Evidence of Validity

CLASS-Bio has been developed and rigorously tested for use in evaluating the novice-to-expert levels of student perceptions about the discipline of biology. The clarity and interpretation of this instrument's statements have been verified through 39 student interviews and expert consensus from 69 PhDs representing 30 institutions. Furthermore, CLASS-Bio is reliable when given to similar students in two different years and can also detect differences among major and nonmajor student populations. CLASS-Bio is unique among epistemological instruments, because it directly compares student responses with those of experts in the field and uses student data to determine categories of student thinking.

While the adaptation of CLASS-Phys for chemistry resulted in a similar instrument, differing primarily with respect to concept-based statements unique to each discipline, CLASS-Bio statements and categories differ more extensively. This difference is likely due to the fact that a higher proportion of statements on CLASS-Bio were either substantially reworded to be applicable to the discipline or are entirely new statements. As a result, the statement composition of CLASS-Bio categories (Table 3) is noticeably different compared with the other instruments and this fact should be considered when comparing responses across CLASS-Phys, -Chem, and -Bio instruments. However, whether statement groupings are dif-

ferent because of a unique pool of statements or because of a difference in how biology students view their discipline is challenging to interpret without further study. While the specific statements vary among the categories, it is important to note that the categories still appear to represent the same fundamental differences between novices and experts in how they view the discipline of biology (e.g., enjoyment of the discipline, likelihood of making connections to the real world, problem-solving strategies, etc.). Furthermore, two independent factor analyses on Fall 2007 and Fall 2008 data resulted in similar categories, lending further evidence that CLASS-Bio categories represent meaningful and consistent aspects of student thinking about biology.

Finally, the fact that some statements are classified into multiple categories has two main implications. First, it reveals interesting aspects of student thinking. For example, statements about the enjoyment of biology are statistically categorized within categories such as Problem-Solving: Effort and Real World Connection. Thus, not only are students' perceptions about how they enjoy the discipline statistically related to their degree of expert-like effort, but helping students to connect course material to real-world applications may in turn help students both enjoy the discipline more and become more expert-like in their problem-solving efforts. Second, having some statements in multiple categories means that outcomes of different categories will be inherently related and may show correlated results, especially among categories

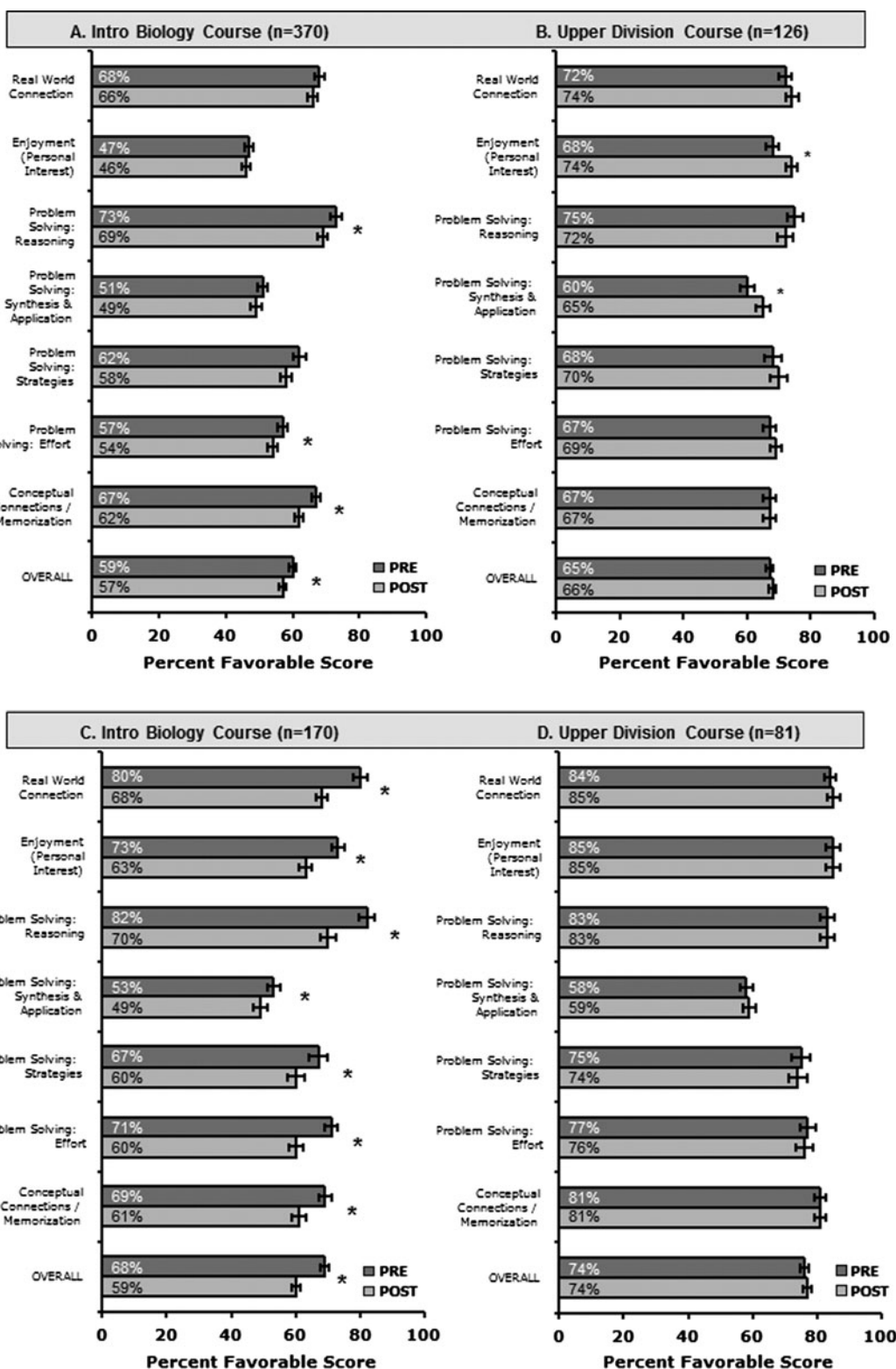


Figure 3. Pre- and postinstruction scores for CLASS-Bio categories in two example curricula with the introductory course of the series and an upper-division course. For all categories in both curricula (A–B and C–D), preinstruction scores in upper-division courses are either comparable to or, in most cases, higher than either entering or exiting scores in each curriculum’s introductory course. While data in both curricula series represent different pools of students between courses (i.e., data do not follow individuals through the curriculum), data across different semesters show consistent patterns of student thinking (see Figure 2). Asterisks denote significant shifts between pre- and postinstruction scores on paired student data within a given category (>2 SEM).

that share multiple statements. However, student responses within similar categories can differ, depending on the course, and thus each category in CLASS-Bio has the potential to offer unique insight into student thinking. For example, the two categories that share the most statements (Problem Solving: Strategies and Problem Solving: Effort) often follow similar patterns of student thinking both pre- and postinstruction. However, in one introductory biology course (Figure 3A), they differ: only Problem Solving: Effort significantly drops postinstruction and preinstruction Problem Solving: Strategies scores are more expert-like than Problem Solving: Effort scores. These data warrant keeping both categories in the analyses of CLASS-Bio data.

Initial Findings Regarding Students Perceptions of Biology

Perception Shifts Following Instruction. The most common way of using CLASS-Bio is to analyze student shifts in perceptions across an instructional period. In both physics and chemistry disciplines, analyses of student perceptual shifts across semesters of introductory courses have shown consistent novice-like shifts following instruction (e.g., Adams *et al.*, 2006; Barbera *et al.*, 2008). We have now demonstrated that this shift also occurs for introductory biology courses. As shown in Figure 2A, introductory biology students from two different institutions shift toward more novice-like thinking over the semester. These shifts are often found across multiple categories of student thinking (Figure 3, A and C), and category shifts on CLASS-Bio are comparable to shifts in other disciplines (Adams *et al.*, 2006; Barbera *et al.*, 2008).

Studies of perceptual shifts in other disciplines have demonstrated that the negative perceptual shifts in introductory courses occur following both traditional instruction and across a variety of nontraditional instructional approaches aimed at improving students' conceptual understanding (Adams *et al.*, 2006). Meanwhile, several groups have now demonstrated that new instructional methods specifically targeting students' perceptions of the discipline can help prevent these negative shifts and even promote shifts toward more expert-like perceptions, as well as lead to increased conceptual understanding of content (Redish *et al.*, 1998; Elby, 2001; Hammer and Elby, 2003; Perkins *et al.*, 2005; Otero and Gray, 2008; Brewe *et al.*, 2009; Moll and Milner-Bolotin, 2009). Thus, improving students' expert-like perceptions appears to require pedagogical techniques that explicitly address epistemological issues.

Even within our preliminary data, we find suggestions that both reversing the reported negative shifts and improving student perceptions are possible over the course of a single semester in a biology course. For example, in the Introductory Biology Course A1 (Figure 2A), student perceptions became significantly more novice-like in 2008, but this novice shift did not exist in 2009. While we did not specifically study the causes of this reversal, after seeing CLASS-Bio results from 2008, this instructor made an explicit effort to include more real-world examples in the 2009 course. In addition, students in one of the upper-division courses had significant expert-like shifts in two specific categories (Enjoyment (Personal Interest) and Problem Solving: Synthesis & Application; Figure 3B). These results suggest that the course's current structure, which includes clicker questions, homework, and frequent

use of health-related applications and examples, promotes expert-like thinking in these specific categories.

Recruitment and Retention. CLASS-Bio can also be used to examine recruitment and retention of students as biology majors. For example, in contrast to the shifts toward novice-like thinking in introductory courses, students in upper-division courses do not as commonly show novice-like shifts (physics: Adams *et al.*, 2006; chemistry: Adams, personal communication). Also, studies in physics have shown that physics majors have more expert-like perceptions than nonmajors when they begin the major, and physics majors show only small perceptual shifts over time until the final year (Perkins *et al.*, 2005; Gray *et al.*, 2008; Gire *et al.*, 2009). Data from CU also indicate that students in upper-division biology courses do not have novice-like shifts and, in some cases, have expert-like shifts (Figures 2B and 3, B and D). Furthermore, among students entering introductory biology courses, majors have higher favorable scores than nonmajors (Figure 1; Knight and Smith, 2010) and similar or higher favorable scores as students entering upper-division courses (Figure 3, B and D). These data suggest that students are not necessarily making large leaps in expert-like perceptions as they move through the major, but rather that students who already enter college with more expert-like perceptions in these fields are being recruited to and retained in the major. More detailed studies using CLASS-Bio to address recruitment and retention of biology majors are currently underway.

Advantages of CLASS-Bio

While other surveys can measure similar perceptions, CLASS-Bio has several advantages. Compared with other instruments about the nature of science, CLASS-Bio: 1) has undergone response validation to ensure that each statement has a single, clear statement interpretation by students, 2) focuses on statements specific to the field of biology, 3) includes multiple areas of novice/expert distinctions in a single tool, and 4) requires relatively minimal effort for students to complete and instructors to analyze. One of CLASS-Bio's greatest strengths is that it is neither course-specific nor a measure of self-efficacy, but rather constitutes perceptions about the biology discipline itself. This gives CLASS-Bio an important advantage over SALG, one of the most common surveys currently used in the field of biology. The SALG instrument asks students to rate their own gains and is thus an inherently subjective measure of students' shifts in perceptions (Seymour *et al.*, 2000). In contrast, CLASS-Bio has students rate their agreement with statements that have been both validated with experts and selected to have a consensus expert response; student responses can thus be directly compared with expert thinking. Therefore, CLASS-Bio provides a more objective measure of changes in student thinking. In addition, when administered with questions designed to elicit student feedback specific to a course (a common practice on postinstruction surveys), one can correlate students' course feedback to students' degree of expert-like perceptions and thus gain even more insight into the nature of student thinking. Combined, these features make CLASS-Bio especially useful when trying to track student perceptions across time, curricula, and institutions.

CONCLUSIONS, INSTRUCTIONS, AND ACCESS TO CLASS-BIO

This paper presents a new and rigorously validated instrument, CLASS-Bio, that can be used in a broad range of undergraduate biology courses to characterize how student perceptions match expert perceptions about how knowledge is structured, where knowledge comes from, and how new knowledge is gained and problems are solved in the field of biology. This instrument can be used to correlate these epistemological perceptions with content knowledge, demographics, and other student characteristics to better understand the impacts of student perceptions on learning and recruitment into a major. Furthermore, this instrument can be used to evaluate impacts of pedagogical reform on students' growth of expertise in scientific thinking.

The online survey and Excel scoring templates to analyze survey results can be found at: <http://CLASS.colorado.edu>. We recommend that the entire survey be used. Even if one is interested in a single aspect of student thinking, the survey was designed and validated to work as a whole, and can be taken online in a short amount of time (less than 10 min). Furthermore, as many aspects of student thinking have been shown to be connected, the full survey will be able to provide the most complete understanding of any single aspect of student thinking.

ACKNOWLEDGMENTS

We thank Wendy Adams, Katherine Perkins, and Matthew McQueen for help with the statistical analysis; Marjorie Frankel and Angela Jardine for their help managing the online survey and summarizing data; and Tamara Kelly and Semira Kassahun with help conducting student interviews.

We thank all CU and UBC faculty who participated in the CLASS-Bio development working groups and all who provided expert opinions on instrument statements. We are very grateful to the 14 faculty teaching biology courses who administered the survey to their students, especially William Adams and Barbara Demmig-Adams.

This project was funded by CU and UBC through the institutions' Science Education Initiative (SEI) and Carl Wieman Science Education Initiative (CWSEI), respectively.

REFERENCES

Adams WK, Perkins KK, Podolefsky NS, Dubson M, Finkelstein ND, Wieman CE (2006). New instrument for measuring student beliefs about physics and learning physics: the Colorado Learning Attitudes about Science Survey. *Phys Rev ST PER 2*, 010101.

Adams, WK, Wieman, CE (2010). Development and validation of instruments to measure learning of expert-like thinking. *Int J Sci Educ 33*, 1289–1312.

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

Anderson DL, Fisher KM, Norman GJ (2002). Development and evaluation of the conceptual inventory of natural selection. *J Res Sci Teach 39*, 952–978.

Barbera J, Adams WK, Wieman CE, Perkins KK (2008). Modifying and validating the Colorado Learning Attitudes about Science Survey for use in chemistry. *J Chem Educ 85*, 1435–1439.

Bowling BV, Acra EE, Wang L, Myers MF, Dean GE, Markle GC, Moskalik CL, Huether CA (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics 178*, 15–22.

Brewe E, Kramer K, O'Brien G (2009). Modeling instruction: positive attitudinal shifts in introductory physics measured with CLASS. *Phys Rev ST PER 5*, 013102.

Crocker L, Algina J (1986). *Introduction to Classical and Modern Test Theory*, Fort Worth, TX: Holt, Rinehart and Winston.

Cronbach LJ, Shavelson RJ (2004). My current thought on coefficient alpha and successor procedures. *Educ Psych Meas 64*, 391–418.

Elby A (2001). Helping physics students learn how to learn. *Am J Phys 69*, S54.

Garvin-Doxas K, Klymkowsky M, Elrod S (2007). Building, using, and maximizing the impact of concept inventories in the biological sciences: report on a National Science Foundation-sponsored conference on the construction of concept inventories in the biological sciences. *CBE Life Sci Educ 6*, 277–282.

Gire E, Jones B, Price E (2009). Characterizing the epistemological development of physics majors. *Phys Rev ST PER 5*, 010103.

Gray KE, Adams WK, Wieman CE, Perkins KK (2008). Students know what physicists believe but they don't agree: a study using the CLASS survey. *Phys Rev ST PER 4*, 020106.

Hammer D (1994). Epistemological beliefs in introductory physics. *Cogn Instr 12*, 151–183.

Hammer D, Elby A (2003). Tapping epistemological resources for learning physics. *J Learn Sci 12*, 53.

Handelsman J, et al. (2004). Scientific teaching. *Science 30*, 521–522.

House JD (1994). Student motivation and achievement in college chemistry. *Int J Instr Media 21*, 1–12.

House JD (1995). Student motivation, previous instructional experience, and prior achievement as predictors of performance in college mathematics. *Int J Instr Media 22*, 157–168.

Howard Hughes Medical Institute (HHMI) (2002). *Beyond Bio101: The Transformation of Undergraduate Biology Education: A Report from the Howard Hughes Medical Institute*, ed. D Jamul. www.hhmi.org/BeyondBio101 (accessed 16 June 2011).

Klymkowsky MW, Garvin-Doxas K, Zeilik M (2003). Bioliteracy and teaching efficacy: what biologists can learn from physicists. *Cell Biol Educ 2*, 155–161.

Knight JK (2010). Biology concept assessment tools: design and use. *Microbiol Aust 31*, 5–8.

Knight JK, Smith MK (2010). Different but equal? How nonmajors and majors approach and learn genetics. *CBE Life Sci Educ 9*, 34–44.

Moll RF, Milner-Bolotin M (2009). The effect of interactive lecture experiments on student academic achievement and attitudes towards physics. *Can J Phys 87*, 917–924.

National Research Council (NRC) (2003). *Bio2010: Transforming Undergraduate Education for Future Research Biologists*, Washington, DC: National Academies Press.

Otero VK, Gray KE (2008). Attitudinal gains across multiple universities issuing the physics and everyday thinking curriculum. *Phys Rev ST PER 4*, 020104.

Panter AT, Swyger KA, Dahlstrom WG, Tanaka JS (1997). Factor analytic approaches to personality item-level data. *J Pers Assess 68*, 561–589.

Perkins KK, Adams WK, Pollock SJ, Finkelstein ND, Wieman CE (2005). Correlating student beliefs with student learning using the Colorado Learning Attitudes about Science Survey. In: *2004 Physics Education Research Conference*, ed. J Marx, P Heron, and S Franklin, American Institute of Physics, 61–64.

Redish EF, Saul JM, Steinberg RN (1998). Student expectations in introductory physics. *Am J Phys* 66, 212.

Seymour E, Wiese D, Hunter A, Daffinrud SM (2000). Creating a better mousetrap: on-line student assessment of their learning gains. In: Proceedings of the National Meeting of the American Chemical Society, San Francisco, CA.

Shi J, Martin JM, Guild NA, Vincens Q, Knight JK (2010). A diagnostic assessment for introductory molecular and cell biology. *CBE Life Sci Educ* 9, 453–461.

Smith MK, Wood WB, Knight JK (2008). The genetics concept assessment: a new concept inventory for gauging student understanding of genetics. *CBE Life Sci Educ* 7, 422–430.

Wood WB (2009). Innovations in teaching undergraduate biology and why we need them. *Annu Rev Cell Dev Biol* 25, 5.1–5.20.

Woodin T, Smith D, Allen D (2009). Transforming undergraduate biology education for all students: an action plan for the twenty-first century. *CBE Life Sci Educ* 8, 271–273.