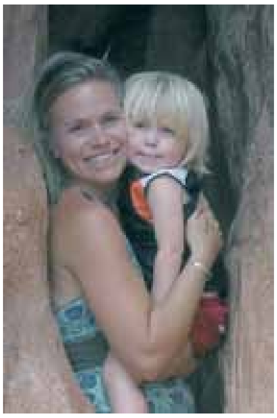
# Biology concept assessment tools: design and use



*Jennifer K Knight*

Department of Molecular Cell and Developmental Biology and the Science Education Initiative, University of Colorado
Boulder, USA
Tel 303 735 1949
Fax 303 492 7744
Email knight@colorado.edu

**Many recent publications have highlighted the need for, and value of, concept assessments (also called inventories) in undergraduate biology education [1-5]. Current interest in such assessments is primarily due to the emergence of a community of science education researchers in biology, who both approach teaching from a scientific perspective and want to measure the potential successes of their teaching reforms [6]. Well-designed, valid and reliable assessment tools that allow instructors to capture student learning of the main concepts of biology are becoming an essential way to inform biology instructors about what students learn in college biology courses. This review summarises the general approaches taken in creating such concept assessment tools and presents some of the ways to effectively use them.**

Concept assessments are sets of multiple-choice questions, often based on common student misconceptions, that are designed to test understanding without relying on memorisation. When the Force Concept Inventory (FCI) [7] was first given to introductory physics students, instructors were surprised by the inability of their students to answer such seemingly easy conceptual questions. While students could use formulas they had memorised to solve complex mathematical problems, they could not answer questions that required an understanding of the concepts without the formulas. These results stimulated many in the physics community to reconsider how they were teaching introductory physics [8]. The results of such efforts, first in physics [9] and more recently in biology [10], have shown that changing the mode of instruction can have a positive impact on student learning. Validated assessment tools like the FCI have proved useful for first identifying problems in conceptual understanding and then benchmarking change by administering the assessment tool again to see whether different pedagogies have improved understanding. As biology educators have embraced the idea that learning biology should be more than simply memorising a collection of facts, a variety of concept assessments have been developed as independent measures of student understanding in particular topics of biology (Tables 1 and 2). As long as an assessment has been designed to address the conceptual nature of particular learning goals or big ideas, it can theoretically be used by anyone with any group of students to measure learning gains or diagnose topics on which students struggle.

Perhaps because biology is such a diverse field, biologists have not necessarily yet agreed on a core suite of 'big ideas' that undergraduate students should learn [3]. This is a particularly difficult problem for introductory biology, since the topics taught in such courses can range from ecology to molecular biology and can include both or neither. It is likely that the best way to ensure the construction of useful concept assessments in the future will be by using the techniques described by D'Avanzo [4], which include faculty development workshops to help teach faculty members about their uses and the involvement of professional societies, which could encourage meaningful assessment and provide avenues for networking. Thus far, the instruments developed have followed a similar but not identical validation process and are intended (as indicated by their respective authors) for slightly different uses. The term 'concept assessment' is used in this review to encompass all the tools that have been developed so far. Some focus on helping to diagnose student pitfalls [11, 12, 19] and others on capturing an overall picture of student understanding [13], but all of them assess student learning.

## Current biology concept assessments

Several recent reviews have listed the currently available instruments for science in general [5] and biology in particular [4]. Table 1 reproduces only the known published biology concept assessments for which the questions are readily available online or from the authors and which have undergone extensive validation and reliability testing. Table 2 reproduces a list of other biology projects that are in development or revision, or are published but without extensive validation.

## Getting the most out of concept assessments:

### Tips for administration

Concept assessments should be given in an environment that encourages students to put effort into the assessment. If the assessment will be used to measure student understanding both before and after instruction (pre-post), it is best to give the assessment on the first day of class prior to any content instruction (pre-test). Students receive participation credit for completing the pre-test; to help ensure that students take the assessment seriously, they can be told that the results will help guide instruction. One should ensure that students do not keep a copy of the assessment, since circulation of the assessment would devalue its future use [14]. The identical assessment is then given at the end of the course as a post-test. One way to administer the post-test is to give it on the next to last day of class, informing students that their answers will be used to help construct a useful review session for the final [15]. An alternative is to give the post-test as part of the final exam [10, 13]. Because students do not know the pre-test questions will be repeated in the final exam, they cannot study specifically for this part of the exam. They will take the questions seriously because they are graded (even 0.5 points per question will still be valued by the students [10]). Ultimately, the most important factor is consistency in administration. If one is interested in comparing learning gains in two different courses, or in the same course over several semesters, the pre- and post-tests should be administered the same way every time. Otherwise, differences in administration will likely impact student performance and make the data sets incomparable [15].

## Using data from concept assessments to inform teaching

Overall improvement in each student's understanding from pre-instruction to post-instruction is called learning gain. To calculate a normalised learning gain ($<g>$), which takes into account the possible gain, the following standard formula is used: $<g> = 100*(\text{post-pre})/(100-\text{pre})$ [9]. This formula can be used for all students in the course unless a student has a negative learning gain, in which case the normalised change formula 100* (post-pre/pre) is recommended instead [16]. The normalised learning gains for all students can then be averaged to give an overall normalised learning gain for a group of students.

For a more fine-grained analysis of what students are learning, one can separate out performance on groups of questions that test individual learning goals and calculate normalised learning gain for each goal [13]. This can be especially useful if one has implemented a new pedagogical technique focused on one topic only. While student interviews are the most informative for exploring student thinking, the most commonly picked distractors (wrong answers) on a well-designed concept assessment can provide a snapshot of student thinking. Normalised learning gains can also be compared among different groups of students in the same course (for example, students who are co-enrolled in

**Table 1. Published, validated biology concept assessments.**

| | |
|---|---|
| Biology Concept Inventory | Garvin-Doxas & Klymkowsky (2008) [19]; http://bioliteracy.net |
| Genetics Concept Assessment | Smith *et al.* (2008) [13] |
| Genetics Literacy Assessment Instrument | Bowling *et al.* (2008) [25] |
| Host-Pathogen Concept Inventory | Marbach-Ad *et al.* (2009) [26] |
| Inventory of Natural Selection | Anderson *et al.* (2002) [27] |

**Table 2. Biology concept assessments under development.**

| | |
|---|---|
| Developmental Biology Concept Assessment | Knight & Wood (2005) [10] |
| Diagnostic Question Clusters (multiple topics) | Wilson *et al.* (2006) [28], Williams *et al.* (2008) [11] |
| Genetics Concept Inventory (GenCI) | Elrod, S. http://bioliteracy.net/CABS.html |
| Introduction to Molecular and Cell Biology Concept Assessment (IMCA) | Shi *et al.* University of Colorado: jia.shi@colorado.edu |
| Molecular Life Sciences Concept Inventory | Howitt *et al.* (2008) [22]; mary.rafter@uq.edu.au |
| Natural selection instrument | Nehm & Reilly (2007) [29] |
| Osmosis and diffusion diagnostic test | Odom & Barrow (1995) [30] |
| Physiology | Michael *et al.* (2009) [31]; Benay, F. *et al.* University of Colorado: francoise.benay@colorado.edu |

another related course in comparison to those not co-enrolled), or among students taking different, but related, courses (for example, non-science majors compared to science majors [17], or two similar courses at two different universities).

## Construction, validation and statistical analyses of concept assessments

The framework for constructing and validating concept assessments has been summarised recently by several authors [2, 5, 15] and generally follows the Standards for Educational and Psychological Testing [18]. Figure 1 briefly outlines these steps, a few of which are highlighted below.

Because one of the goals of concept assessments is to reveal what students are thinking when they have an incorrect or alternative conception of how something works, it is critical to capture student thinking at the outset of generating an assessment. Once a set of concepts has been agreed upon by faculty, open-ended, 'think-aloud' student interviews, in which students explain their answer to an open-ended question, can be used as the starting point for building multiple-choice questions [13, 14, 19]. Such student interviews are also valuable for hearing how students use natural language to describe biology concepts. This in turn allows the assessment to be worded with less jargon, ensuring that students can understand the question and that one is not testing only knowledge of vocabulary.

After a pilot assessment has been written and administered to students, a second round of student interviews is required for validation: in these interviews, students select the correct answer from the choices and explain their reasoning. Questions for which students pick the correct answer but supply incorrect reasoning must be reworded. In addition, each distractor should be chosen as a correct answer by some students in order to be considered a true distractor that captures student thinking. Finally, each question should address only one concept; the answers should avoid words like none, always and never; and the answers should be of equal lengths and complexity [5]. Once questions are constructed and tested through student interviews, experts should be asked to review the assessment. One method is to ask experts to take the assessment and rate each question for clarity, scientific accuracy and how well the question addresses the learning goal it was written to assess [13]. Questions that are not highly rated by experts should then be rewritten and revalidated.

Three statistical measures are common for collecting evidence of validity: item difficulty, item discrimination and reliability of the instrument [14, 20]. Item difficulty (P) measures the percentage of students who answer a question correctly; thus one expects to see this value increase from the pre-test to the post-test for most questions (high P=high % of students answering question correctly; low P=low % of students answering question correctly). However, some questions remain difficult and thus will have a low P value in both pre- and post-tests. When combined with item discrimination, D (how well a question discriminates between students who have performed well on the assessment overall vs students who have performed poorly on the assessment overall), the two values can provide insight into student understanding. For example, items that begin with a high D value on the pre-test and end with a similarly high D value but low P value on the post-test are concepts that remain challenging for all but the strongest students in a course [13]. In addition, the overall difficulty of the assessment should be considered. If the average pre-test performance on a concept assessment is above 70%, the overall possible gain from pre-test to post-test is smaller than if the average performance is below 50%. Thus, an average pre-test performance between 25 and 40% (25% would be the score expected if students were guessing, given an average of four answers for each question) is common [9, 13, 14].

**Figure 1. Steps for constructing biology concept assessments*.**

1. Review literature on common misconceptions on subtopic of biology.
2. Interview faculty to develop a set of learning goals describing the concepts central to a particular subtopic of biology.
3. Use 'think aloud' interviews or analysis of student essays on these concepts to gather information about student thinking.
4. Generate and administer a pilot assessment based on known and perceived misconceptions using student–provided distractors and natural language.
5. Validate and revise the questions through student interviews and expert review.
6. Administer the assessment to many students (preferably >300, in several courses at several institutions).
7. Evaluate the assessment using statistical analyses to measure item difficulty, item discrimination and reliability.
8. Repeat steps 5-7 as necessary.

*adapted from [13,15]

Finally, for measuring reliability, a commonly used test is the test-retest approach [20, 21], where the assessment is given to students taking the same course in consecutive semesters and the coefficient of stability is then calculated between the two sets of pretest results. The internal consistency test (Cronbach's $\alpha$) is favoured by designers of standardised tests; however, concept assessments comprise questions that test different concepts and thus performance on individual concepts need not be correlated [15].

## Challenges of biology concept assessments

As mentioned earlier, widely applicable concept assessments in biology are challenging to build for at least two reasons. Firstly, different biology content is emphasised differently in each department at each institution and secondly, biologists have not yet agreed on what concepts are most important to assess even within sub-disciplines. Because most instructors want to use an assessment that is relevant to the content taught in their courses, they may be tempted to use only subsets of questions from published concept assessments, choosing the questions that are most valuable to them. Some assessments are designed to be offered this way (e.g. the Biology Concept Inventory [19] and the Molecular Life Sciences Inventory [22]). However, others suggest that an assessment is only valid if always administered with the same questions in the same order [23].

Several additional issues remain for biology faculty to consider. Should all concept assessments rely on the multiple choice format? Can critical thinking or logical skills be assessed with multiple choice questions, or should written assessments with rubrics be developed and validated? How important is the mastery of process skills such as hypothesis testing and experimental design and should such skills be measured within concept assessments intended for lecture courses, or through separate assessments designed for use in laboratory classes [24]? Whatever the directions chosen by biology educators, concept assessments will continue to provide valuable feedback to instructors about student thinking and learning.

## References


1. Pellegrino, J.W. *et al.* (eds). (2001) *Knowing what students know: the science and design of educational assessment*. NRC, National Academy Press, Washington, DC.
2. Richardson, J. (2005) Concept inventories: tools for uncovering STEM students' misconceptions. In: *Invention and Impact: Building Excellence in Undergraduate Science, Technology, Engineering and Mathematics (STEM) Education*. American Association for the Advancement of Science, New York.
3. Garvin-Doxas, K. *et al.* (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.
4. D'Avanzo, C. (2008) Biology concept inventories: Overview, status and next steps. *BioScience* 58, 1079-1085.
5. Libarkin, J.C. (2008) *Concept Inventories in Higher Education Science.* National Research Council Promising Practices in Undergraduate STEM Education Workshop 2 ,Washington, DC, 13-14 October 2008.
6. Handelsman, J. *et al.* (2004) Scientific teaching. *Science* 304, 521–522.
7. Hestenes, D. *et al.* (1992) Force Concept Inventory. *Phys. Teach.* 30, 141–158.
8. Mazur, E. (1997) *Peer Instruction: A User's Manual*. Prentice Hall, Saddle River.
9. Hake R.R. (1998) Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *Am. J. Phys.* 66, 64-74.
10. Knight, J.K., & Wood, W.B. (2005) Teaching more by lecturing less. *CBE Life Sci. Educ*. 4, 298-310.
11. Williams, K. *et al.* (2008) *Using diagnostic test items to assess conceptual understanding of basic biology ideas: A plan for programmatic assessment.* Conceptual Assessment in Biology Conference II; 3–8 Jan., Asilomar, CA. At: http://bioliteracy.net/CABS.html
12. Klymkowsky, M.W. *et al.* (2003) Bioliteracy and Teaching Efficacy: What Biologists can learn from Physicists. *CBE Life Sci. Educ.* 2, 155-161.
13. Smith, M.K. *et al.* (2008) The Genetics Concept Assessment: a new concept inventory for gauging student understanding of genetics. *CBE Life Sci. Educ.* 7, 422-430.
14. Ding, L. *et al*. (2006) Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Phys. Revs. Special Topics – Phys. Educ. Res.* 2, 010105.
15. Adams, W., & Wieman, C.E. (submitted) Development and validation of instruments to measure learning of expert-like thinking. *J Sci. Educ.*
16. Marx, J. & Cummings, K. (2007) Normalized Change. *Am. J. Phys.* 1, 87-91.
17. Knight, J.K. & Smith, M.K (2010). Different but Equal? How non-majors and majors approach and learn genetics. *CBE Life Sci. Educ.* 9, 34–44.
18. *Standards for Educational and Psychological Testing* (AERA, APA, NCME) (1999).
19. Garvin-Doxas, K. & Klymkowsky, M.W. (2008) Understanding Randomness and its Impact on Student Learning: Lessons Learned from Building the Biology Concept Inventory (BCI). *CBE Life Sci. Educ.* 7, 227-233.
20. Adams, W.K. *et al.* (2006) New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Phys. Revs. Special Topics – Phys. Educ. Res.* 2, 010101.
21. Crocker, L. & Algina, J. (1986) *Introduction to Classical and Modern Test Theory*. Holt, Rinehart and Winston, New York.
22. Howitt, S. *et al.* (2008) A Concept Inventory for Molecular Life Sciences: How will it help your teaching practice? *Aust. Biochem.* 39, 14-17.
23. Maloney, D.P. *et al.* (2001) Surveying students' conceptual knowledge of electricity and magnetism. *Am. J. Phys. Phys. Educ. Res. Suppl.* 69, S12-S23.
24. Brickman, P. *et al.* (2009) Effects of inquiry-based learning on students' science literacy skills and confidence. *Int. J Schol. Teach. Learn. 3*, 1-22.
25. Bowling, B. *et al.* (2008) Development and evaluation of a Genetics Literacy Assessment Instrument for undergraduates. *Genetics* 178, 15-22.
26. Marbach-Ad, G. *et al.* (2009) Assessing student understanding of host pathogen interactions using a concept inventory. *JMBE* 10, 43-50.
27. Anderson, D.L. *et al.* (2002) Development and evaluation of the conceptual inventory natural selection. *J. Res. Sci. Teach*. 39, 952-978.
28. Wilson, C.D. *et al.* (2006) Assessing students' ability to trace matter in dynamic systems in cell biology. *CBE Life Sci. Educ. 5*, 323-331.
29. Nehm, R.H. & Reilly, L. (2007) Biology majors' knowledge and misconceptions of natural selection. *BioScience* 57, 263-272
30. Odom, A. L. & Barrow, L. H. (1995) Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction. *J. Res. Sci. Teaching* 32, 45-61.
31. Michael, J. *et al.* (2009) The "core principles" of physiology: what should students understand? *Advan. Physiol. Educ.* 33, 10-16.


## Biography

Jenny Knight is a senior instructor in the Department of Molecular Cell and Developmental Biology at the University of Colorado, Boulder, where she has taught for 10 years. With a background in neuroscience and developmental genetics, she has focused on biology education research since 2004 and has been the coordinator of the MCDB division of the Science Education Initiative (http://www.colorado.edu/sei/) since 2006. She wishes to thank Jia Shi, Michelle Smith and Bill Wood for their integral roles in this work.