

Conceptual Assessment in Biology, III

May 17-20, San Diego, CA

Response from Kelly Myer Polacek

June 15, 2010

Contents

Describe your status, perhaps as a newcomer to the field.	1
Summarize your reactions to the meeting.....	2
Issue 1: The need for CABs to be widely available	3
Issue 2: A centralized CAB website.....	4
Issue 3: Differing and competing CABs	5
Possible Improvements	7
References.....	8

Describe your status, perhaps as a newcomer to the field.

I hold two Master's degrees—one in the life sciences with an emphasis in science education, and Master of Library Science with an emphasis in science literacy—and I am a member of the Board of Editors in the Life Sciences. I have spent the last 10 years studying the development of critical thinking among college science students. Some of this work occurred in formal academic coursework. Other perspectives developed as I integrated cognitive development and critical thinking frameworks within science literary courses. Additionally, I collaborated with several colleagues in publications on the science of teaching and learning, always maintaining a perspective that students need and deserve to have deep conceptual understanding of the concepts we as scientists deem important.

I learned about concept inventories only 2 years ago when Susan Elrod invited me to help develop a concept inventory in genetics. We experienced the typical challenges faced when writing conceptual questions, primarily differentiating between what is and is not a concept, and designing jargon-free questions about genes. As our work developed we began to see that despite

our diligent efforts, the conceptual questions we designed would never be wholly inclusive, wholly jargon-free, or satisfactory to any given expert.

We paused to reflect on our ultimate goal: That students gain deep conceptual understanding of fundamental scientific processes. To accomplish this, we acknowledged the need for informed, reflective instructors. As such, we shifted our focus from writing an impossibly-comprehensive genetics concept inventory and instead began a series of question clusters faculty can use to diagnose misconceptions held about certain topics in genetics (D'Avanzo et al., 2008; Elrod & Polacek, 2009). By refocusing our efforts, we then generated conceptual questions on individual topics and began accompanying those with instructor-specific information including explanations from the literature as to why students might display misconceptions, specific teaching strategies for classroom implementation, and follow-up assessments for instructor use to determine learning gains.

Given that such assessment could inform my own instruction, I redesigned my upper-division physiology course to include two forms of pre-/post-assessment: 1) a small number of conceptual questions in physiology collected from several sources (since a physiology CAB does not currently exist); and I correlated those results with data from 2) concept maps students generated in answer to the question “How do organisms adjust to changing environments?” It can be difficult for individual faculty to adopt alternative teaching strategies without support from internal colleagues; as such, I am reporting my experiences in the form of a guide to assist faculty in the implementation of conceptual assessment before, during, and after a course, and how supervisors can include such informed, self-evaluation in merit review of faculty (Polacek, in preparation). Indeed, for a community that typically relies on statistical data from controlled interventions to evaluate phenomena, teaching faculty in the sciences need more than the uninformed opinions found in student evaluations to evaluate their own teaching effectiveness. CABs could be the first step toward teacher evaluation reform—providing the needed raw data to demonstrate learning occurs in our college classrooms.

Summarize your reactions to the meeting.

Two issues prevented us from making great strides: 1) whether or not CABs, either partial or in their entirety, should be publicly available; and if so, 2) the logistics of how to make them available on a centralized CAB website. A third issue that permeated the meeting but was never overtly discussed was that the CABs authored by the participants differed significantly in many of the areas addressed in “Box A. CAB III Outcomes” below. This last issue was of significant importance to me, both as a relative newcomer to the field of conceptual assessment, and as a veteran of science education. If we are promoting pedagogical reform in higher education, what role might conceptual assessments hold and how will their similarities and differences impact that potential change?

Issue 1: The need for CABs to be widely available

A significant number of scholars feel that CABs should be kept password protected and there is similar interest in making CABs widely available for instructor use. In a recent listserv posting, Hake writes, “In my opinion, such hard-won Diagnostic Tests that cover important parts of common introductory courses are national assets whose confidentiality should be as well protected as the MCAT (Medical College Admission Test). Otherwise the test questions may migrate to student files and thereby undermine education research that relies upon the validity of such tests” (Hake, 2010). I see at least two flaws with this logic. The first is that by definition, conceptual questions are to be jargon-free; that is, it should be difficult if not impossible for students to look up an answer in a textbook. For example, after a short paragraph describing the dynamics of Venezuelan guppy populations, the following question appears on the Conceptual Inventory of Natural Selection (Anderson, Fisher, & Norman, 2002), “In guppy populations, what are the primary changes that occur gradually over time?” Multiple choice answers include the terms population, traits, offspring, mutations, and environment. Reading the pages of a text on any of those terms will not answer the question about gradual changes in guppy populations over time. In other words, if the test is generated appropriately, it is unlikely students will be able to or interested in finding the right answer. The second flaw with this claim is comparing current biology concept inventories to standardized graduate entrance exams. Approximately 50% of the MCAT exam questions are knowledge or comprehension, the lowest ranking of Bloom’s taxonomy; furthermore, the MCAT offers no questions at the synthesis or evaluation levels (Zheng, Lawhorn, Lumley, & Freeman, 2008). Considering that more than half of the test can be memorized, it makes sense to fiercely guard its questions.

Other faculty are concerned that students who gain access to the CAB will memorize answers thus influencing authentic measures of student understanding. The Force Concept Inventory has been available in the published literature since 1992 (Hestenes, Wells, & Swackhamer). Nevertheless, in his analysis of teaching methods using FCI scores from 6,000 students, Hake notes no difference in baseline scores over several years despite the test’s availability (Hake, 1998). And, since CABs are used as diagnostic tools and not summative assessment, there is no intrinsic need for students to memorize answers to a test that does not count towards their course grade. (Indeed, if the ultimate goal of CAB use is to increase student understanding, then those diligent students who insist on “preparing” for CAB assessment will have accomplished our work on their own!)

Finally, there is some concern that faculty will gain access to CABs without the benefit of instruction on how to properly implement the test, assess the outcomes, and adjust teaching approaches to best address the misconceptions revealed by the diagnosis. Some of this opinion may come from the expert blind spot held by CAB developers (Nathan & Petrosino, 2003). As experts on the nuances of CABs they fail to recognize that there are simply not enough tools available to the average science instructor interested in informing their own instruction. It could be argued that any effort (i.e. the complete, partial, or modified adoption of a validated CAB) on the part of instructors to improve teaching should be encouraged. There are currently several

models for faculty development interested in improving conceptual teaching and learning in their classrooms (D'Avanzo et al., 2008); see also the NSF-funded *Thinking Like a Biologist* website (www.biodqc.org).

Many believe that only when conceptual assessment instruments are widely available will they begin to impact course instruction, curricular design, and student learning. The success of open source practice in information technology has been successfully applied to the life sciences, including the Human Genome Project (U.S. Department of Energy & National Institutes of Health, 2010) which decoded the human genome years ahead of schedule. There is something to be said for open source as the underdog's hope for fast progress and widespread implementation—India's Open Source Drug Discovery project has made tuberculosis a focus in the war on disease. With more than a third of the world's population suffering from TB and only six drugs on the market to treat it, OSDD created a portal to the largest database on *Mycobacterium tuberculosis* and currently has more than 700 participants contributing to its “open-lab notebook” (Menon, 2009). These are great examples Carl Weiman's proposal to apply scientific thinking to science education research (2007). If science has success using open source strategies to decode genomes and develop drugs, why not apply similar open source strategies to develop pedagogical and assessment techniques that will fundamentally reform higher education? “Open source-ness” is perhaps the most important benefit of a centralized CAB website—not only would such a website serve as a repository of CABs, but it could include places and spaces for instructors and researchers to play, collaborate, and develop useful instruments specific to instructional need. See “A centralized CAB website” below.

Issue 2: A centralized CAB website

The time is right for creating a web presence for educational research in teaching and learning of conceptual understanding. There are several existing hosts that could collaborate in such a dissemination site (see “Possible CAB Websites” box). We should define the characteristics of a central dissemination website and focus efforts on obtaining funds necessary for development of such a website. Regardless of the location of this site, it is important to note that distribution and dissemination alone are not enough to cause the kind of informed and reformed instruction necessary to meet the need for increased science literacy in our students. This adds weight to our mission—we must think about what is important to include regarding dissemination of conceptual assessments in biology, support for development of CABs, and support for users of CABs. We must provide ongoing support, a robust pedagogical framework, and

Possible CAB websites



Inqueri
Ohio State University
www.inqueri.org



Thinking Like a Biologist
Michigan State University
www.biodqc.org



ciHUB
Purdue University
<http://cihub.org>

regular workshops. It would be valuable to create a series of short videos illustrating how to use pre- and post- assessments to inform teaching, and to illustrate a variety of approaches to interactive teaching in large lecture classes. These should be readily accessible from the front page of the CAB web site. If we could develop a plan for these, including identifying different strategies and people who are effective in using them, this could be incorporated into the grant proposal. The website could also offer annotations to each conceptual question, since it is all too easy to accept (rather than question) success (in teaching). Annotation also enables instructors (and students) to learn something specific about what students are thinking, assuming, remembering, etc.

Issue 3: Differing and competing CABs

To date, there are 4 inventories that focus genetics, each rationalized by some small difference from the others (Bowling et al., 2008; Garvin-Doxas & Klymkowsky, 2008; Smith, Wood, & Knight, 2008; Tsui & Treagust, 2010). Furthermore, by comparing these genetics CABs, it became evident that despite the outward agreements of the group, it seems in practice scholars do not agree on the characteristics of a CAB question, the purpose of a single CAB question or a collection of reliable and validated questions, and the reason for developing a CAB in the first place. Among the four inventories, Bowling et al.'s (2008) *Genetics Literacy Assessment Instrument* best adheres to the original (Treagust? Hestenes?) instructions for the design of conceptual assessment questions: They are jargon-free, reliable and valid, and address the concepts identified as essential by an American Society of Human Genetics committee whose members included specialists in undergraduate biology education, clinical genetics, nursing, genetics education, and had more than 100 years of cumulative teaching experience committee (Hott et al., 2002).

Smith et al.'s (2008) *Genetics Concept Assessment* presents a number of concerns when it comes to essential concept inventory criteria. First, consider the cognitive load required to correctly answer the following question: “Two different genes are located on the same chromosomal pair in rabbits. A particular female rabbit is heterozygous for alleles of both these genes, with the alleles arranged as shown in the diagram to the right. Scientists know that the two genes are on the same chromosome, but do not know their exact position, as indicated by the dashed line. Suppose this female mates with a male rabbit in which the same chromosome pair looks like [the image at right]. How likely is it that this pair of rabbits would have offspring with a chromosome pair that looks like [the image at right]?” The question is filled with jargon, requires students to process unfamiliar diagrams with straight and dashed lines, as well as understand the differences among alleles, genes, and chromosomes, terms unfortunately proven to be consistently difficult for students to understand (Lewis & Wood-Robinson, 2000; Lewis & Kattmann, 2004; Pashley, 1994). Furthermore, questions for the GCA were designed to meet course goals; the title of their Table 2 clearly delineates this: “Table 2. Course learning goals for the [University of Colorado] majors genetics course and corresponding questions on the GCA.” In fact, the topics listed there are the ones sent to experts in order to validate the test. There is

something flawed in a development scheme that starts with a list of learning goals, generates questions towards those goals, and then validates the test as a general genetics concept inventory based solely on one university's interpretation. In other words, their concept inventory is ideal for assessing learning at their institution, it can be used to document change as supported by external funding, and as such, there is a need to protect the contents of the *GCA*. Unfortunately, the *GCA* cannot be both protected and used by other instructors/institutions simultaneously.

I raise these concerns to bring attention to the fact that some education scholars developing CABs have ignored their own expert blind spots. There is no need for majors, non-majors, two-tiered, and random-based CABs—*most faculty conduct no kind of formative assessment before, during, or after instruction*. Furthermore, with average or above average student evaluation scores, many see no need to authentically evaluate student performance: They are receiving satisfactory reports from students—with instruction meriting little of the tenure process at many institutions, what motivation do they have to change?

The need for informed and reformed instruction is so great, it would be a travesty if competing CABs, differing implementation instructions, and conflicting assessment interpretations prevented a budding reformed instructor from considering including formative assessment in their own courses. The reasons behind CAB development are rooted in the desire for good teaching and deep learning; it is important we maximize that intent by making CABs as applicable within the discipline and across institutions: open and available for collaboration.

Box A. CAB III Outcomes

Characteristics of CABs

- Focus on a single, fundamental, BIG idea
- Assess conceptual understanding at higher cognitive levels than recall
- Are formative assessments; not for sorting students or assign grades
- Can diagnose student misconceptions
- Follow best practices in item design
- Are jargon-free
- Provide valid and reliable results
- Are broadly applicable across institutions

How CABs are developed

- Identify goals
- Develop specification table
- Create items of varying difficulty
- Test CAB with students.
- Replace distracters that are not attractive to students; aim for 30-70% overall.
- Test CAB with experts; aim for 100% agreement.
- Interview students to assess interpretations of item headers and responses
- Determine validity and reliability.

Types of CABs

- Type I – Valid, reliable, intact
- Type II – partial, clustered, DQCs

CAB variations

- May have closed (MC, TF) or open responses
- Distracters may be based on misconceptions
- Some have two-tiered questions
- Can include scenarios, diagrams, schematics
- Some have more than one item assessing a single concept
- Some require students to explain responses

Possible Improvements

To date, approximately 25 concept inventories in biology have been validated and published, and generally fall into two categories. Type I CABs have a broad focus on an entire instructional area such as biology (Garvin-Doxas & Klymkowsky, 2008) or genetics (Bowling et al., 2008). These are generally referred to as Concept Inventories (CI). When properly created, they are reliable and valid in their intact form. Type II CABs center on a particular big idea within biology such as natural selection (Anderson et al., 2002), or host-pathogen interactions (Marbach-Ad et al., 2009). These are sometimes referred to as concept inventories, but occasionally are clustered on a topic and referred to as Diagnostic Question Clusters (D'Avanzo et al., 2008). CABs can be used in a variety of ways: 1) to assess students' prior knowledge to inform instruction (formative assessment); 2) assess students' knowledge pre- and post-instruction to determine learning gains or losses; 3) identify the frequency of selected commonsense ideas or misconceptions among the student population to inform instruction; 4) may be used by students to assess own knowledge; and 5) can be used to determine effectiveness of particular teaching strategies.

The CAB III meeting was an ideal place to have a “thesis defense” of sorts on the CABs currently available and on questions currently under development. For example, working groups agreed upon a variety of characteristics of CABs (see “Box A. CAB III Outcomes”) which don’t necessarily align with the original works of Hestenes et al. (1992) and Treagust (1988) or the summative report by Libarkin (2008). It would have been useful to poll participants as to their opinions on how a particular question met the characteristics defined by the group versus those fundamental papers. For example, *pick two CAB questions (published or unpublished) and matrix them across CAB III criteria vs. Libarkin (2008) criteria. Describe your findings.* Such an exercise could have elucidated differences in interpretations and applications, and the usefulness of such differences.

Additionally, entire CABs could have been peer-reviewed by the larger group. One participant is working on a concept inventory about osmosis and diffusion. Are there really 20 high-quality, genuine conceptual questions about osmosis and diffusion? What about the fact that a concept inventory in osmosis and diffusion already exists (Odom & Barrow, 1995)? The widely-used Biology Concept Inventory (BCI) isn’t really about biology; at last viewing, 66% of its questions address genetics and evolution, leaving only 33% of questions to address the concepts of cells, energetics, plant and animal physiology and behavior, and ecological systems. Unfortunately, the BCI includes no questions directly about plants, animals, or ecology.

Clearly we have much to learn about the design, implementation, and distribution of conceptual assessment tools in the life sciences. Future collaborative meetings will help address some of these concerns, and perhaps raise heretofore unthought of applications and solutions for these valuable tools

References

- Anderson, Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, 39(10), 952-978. doi:10.1002/tea.10053
- Bowling, B. V., Acra, E. E., Wang, L., Myers, M. F., Dean, G. E., Markle, G. C., Moskalik, C. L., et al. (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. *Genetics*, 178(1), 15.
- D'Avanzo, C., Morris, D., Anderson, A., Griffith, A., Williams, K., & Stamp, N. (2008). Diagnostic question clusters to improve student reasoning and understanding in general biology courses: Faculty development component. Presented at the CAB II Meeting, Asilomar, CA. Retrieved from <http://bioliteracy.net/manuscripts08.pdf>
- Elrod, S., & Polacek, K. (2009). Diagnostic question clusters: Helping faculty diagnose and treat misconceptions in genetics.
- 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(2), 227-233. doi:10.1187/cbe.07-08-0063
- Hake. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66, 64-74.
- Hake, R. (2010, June 4). Why Password Protection of Concept Tests is Critical: A Response to Klymkowsky. *LISTSERV 16.0 - AERA-L Archives*. Retrieved July 2, 2010, from <http://listserv.aera.net/scripts/wa.exe?A2=ind1006&L=AERA-L&F=&S=&X=564DD601BDCE4504B9&Y=rrhake%40earthlink.net&P=255>
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3), 141-158.
- Hott, A. M., Huether, C. A., McINERNEY, J. D., Christianson, C., Fowler, R., Bender, H., Jenkins, J., et al. (2002). Genetics content in introductory biology courses for non-science majors: theory and practice. *BioScience*, 52(11), 1024-1035.
- Lewis, & Kattmann, U. (2004). Traits, genes, particles and information: Re-visiting students' understandings of genetics. *International Journal of Science Education*, 26(2), 195-206. doi:10.1080/0950069032000072782
- Lewis, & Wood-Robinson, C. (2000). Genes, chromosomes, cell division and inheritance-do students see any relationship? *International Journal of Science Education*, 22(2), 177-195.
- Libarkin, J. (2008). Concept inventories in higher education science. Presented at the Undergraduate STEM Education Workshop II, Washington DC.
- Marbach-Ad, G., Briken, V., El-Sayed, N., Frauwirth, K., Fredericksen, B., Hutchenson, S., Gao, L., et al. (2009). Assessing student understanding of host pathogen interactions using a

- concept inventory | marbach-ad | journal of microbiology & biology education. *Journal of Microbiology & Biology Education*, 10. Retrieved from http://jmbe.asm.org.ezproxy.lib.calpoly.edu:2048/index.php/jmbe/article/viewArticle/98/html_5
- Menon, S. (2009, March 1). Science 2.0 is here as CSIR resorts to open source drug research for TB. *Business Standard*. Retrieved from <http://www.business-standard.com/>
- Nathan, M. J., & Petrosino, A. (2003). Expert blind spot among preservice teachers. *American Educational Research Journal*, 40(4), 905-928. doi:10.3102/00028312040004905
- Odom, A. L., & Barrow, L. H. (1995). Development and application of a 2-tier diagnostic-test measuring college biology students understanding of diffusion and osmosis after a course of instruction. *Journal of Research in Science Teaching*, 32(1), 45-61.
- Pashley, M. (1994). A-level students: Their problems with gene and allele. *Journal of Biological Education*, 28(2), 120-126.
- Polacek, K. (n.d.). Alternative inclusions in your teaching portfolio: Evidence beyond student evaluations of teaching performance.
- Smith, Wood, W., & Knight, J. (2008). The Genetics Concept Assessment: A New Concept Inventory for Gauging Student Understanding of Genetics. *CBE-Life Sciences Education*, 7(4), 422-430.
- Treagust, D. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 1464-5289.
- Tsui, C., & Treagust, D. (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *International Journal of Science Education*, 32(8), 1073-1098. doi:10.1080/09500690902951429
- U.S. Department of Energy, & National Institutes of Health. (2010, June 23). Human Genome Project Information. Retrieved from http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml
- Wieman, C. (2007). Why not try a scientific approach to science education? *Change: The Magazine of Higher Learning*, 39(5), 9-15. doi:10.3200/CHNG.39.5.9-15
- Zheng, A. Y., Lawhorn, J. K., Lumley, T., & Freeman, S. (2008). Application of Bloom's taxonomy debunks the "MCAT Myth". *Science*, 319(5862), 414-415. doi:10.1126/science.1147852