The Biology Concept Inventory (BCI) Project: Status and Issues

Kathy Garvin-Doxas and Michael W. Klymkowsky
University of Colorado, Boulder

Thought paper prepared for the Conceptual Assessment In Biology (CAB II) Conference

a. What are the biological "big ideas" that are related to and being informed by your work?
b. Describe the concept inventory or other assessment tool you are developing, your results, and plans for use by and dissemination to a broader community.
c. How are the results of your inventory informing and improving student learning in biology course(s)?
d. What are your challenges? What help could you use from others?

Below we briefly explore aspects of the Biology Concept Inventory (BCI) project as they relate to the questions posed in the RFP for this meeting. For discussion of the previous CAB meeting, please see Garvin-Doxas, Klymkowsky, and Elrod (2007; in press). For further discussion of these and related issues, please see:
Garvin-Doxas, Doxas, and Klymkowsky (2007) *Ed's Tools: A Web-based Software Toolset for Accelerated Concept Inventory Construction.* In: Deeds D, editor; Washington DC, October 19-21, 2006.
Klymkowsky and Garvin-Doxas (under review) *Understanding Randomness and its Impact on Student Learning: Lessons Learned from Building the Biology Concept Inventory*

"Big Ideas" measured by the BCI

The Biology Concept Inventory (BCI) is a multiple-choice formatted instrument that measures several critical "big ideas" in the biological sciences. These include the idea of randomness in both molecular motion and interaction, as well as genetic drift; the logic of genetic interactions; the use of energy in biological systems; basic properties of key molecular; the ways plants and animals use energy; the role of controls in scientific experiments; and simple aspects of evolutionary theory. Our research and development has shown that the particular mental models that students hold in these areas to be the sort of root conceptual understanding that students require in order to fully understand several subdisciplinary fields at once.

At this point, our research has progressed the furthest in the areas of student understanding of random processes (Klymkowsky & Garvin-Doxas, underreview). We discovered that students believe that random processes are inefficient, while biological systems are inherently very efficient. As a result of this foundational belief, they are quick to propose their own rational explanations for various processes (e.g., processes that range from diffusion to evolution - a variation of the historical argument by design). These rational explanations almost always make recourse to a driver, such as natural selection in evolution or concentration gradients in molecular biology, with the process taking place only when the driver is present. The concept of underlying random processes that occur all the time and giving rise to emergent behavior is almost totally absent in students. Even students who have had advanced or college physics, and can discuss diffusion correctly in that context, cannot make the transfer to biological processes, and passing through multiple conventional biology courses appears to have little effect on their underlying beliefs. Students' faulty mental model also impacts their ability to truly master anything related to research on evolving populations (something covered only tangentially in the current set of BCI questions). This understanding of student thought emerged during our research (the BCI was not designed to replace the Natural Selection Concept Inventory developed by Anderson, et al, 2002), and so we describe how we came to understand students' misconceptions about the big idea encompassed by "randomness" and its relationship to evolutionary as well as molecular processes.

As part of the research involved in the development of the BCI, we administered a large and varied array of open-ended essay questions and asked them to employ at least 100 words in their responses. This data was collected from students taking undergraduate biology courses at several different institutions. We employed this initial step in our methodology in order to collect a wide array of students' natural language, as well as an effort to begin to understand the mental models students hold in their heads. This is an important point: rather than examining data in terms of *a priori* assumptions or hypothesis testing, we are involved in an exploration, through the iterative and inductive process underlying BCI development, to determine "what is there?" with respect to student thinking about broad topical areas. A level of understanding that is commonly overlooked in both instruction and by conventional tests.

Initial data collection and analysis seeks to identify areas where students' explanations and understanding appear to be "fuzzy." Student responses to three different essay questions during our first round of data collection indicated that they experience some sort of challenge when it comes to conceptual level understanding of natural selection. None of the three questions used focuses on natural selection directly, but rather examine student understanding of evolutionary processes in a general way. Content analysis of this data found that it was common for students to write about natural selection in contradictory ways and that there were several recurring patterns among their responses and across the three questions. Unfortunately, the precise nature of students' difficulties were unclear in terms of their responses to the essay questions. The recurring patterns indicated that there was *something* students were *probably* holding misconceptions about, but the "big idea" that explained the meaning of their responses remained hidden. This indicated the need to conduct thematic interviews on the subject in order to discover whether or not this was an area that needed to be explored further. Results of these interviews indicated that the *something* we could not quite identify in responses to essay questions was the idea of random processes as they relate to evoluation. A recent paper by the population geneticist Michael Lynch (2007) suggests that a number of professional biologists have similar pre-/ misconceptions.

By the time we completed our first round of interviews, we realized that students experience difficulties (at the conceptual level) with random processes in other areas of the biological sciences as well, and we began to develop additional questions for the BCI to explore their understanding of randomness in these other contexts. For example, essay responses to the question, "What is diffusion and why does it occur?" were consistent with some of the ways in which students characterized mutations (particularly during interviews), in the sense that one thing they tend to fail to mention about mutations is that they occur all the time, and (essentially) randomly. In both cases, while students were busy explaining about, and listing characteristics of, diffusion and mutations they consistently failed to include that either held any random component. It was not so much a matter of what they did say, but about what they consistently left out. There is no apparent appreciation displayed that random processes can give rise to emergent behavior. The overall result of these findings is a series of BCI questions exploring students' understanding of randomness across multiple contexts in biology. We have a complete paper exploring these results and their implications in detail that is forthcomming (Klymkowsky& Garvin-Doxas, under review).

Another "big idea" covered by the BCI is students' tendency to apply a geometric rather than an energy model to the understanding of molecular affinity. For any number of reasons (including the pictures used in textbooks), students often become "stuck" on a geometric model for molecular affinity. In other words, they have a tendency to believe molecular affinity works much like a jigsaw puzzle. The research we conducted as part of the development of the BCI demonstrates that even when they learn to respond correctly to many of the traditional questions about molecular affinity, the mental model that persists is the geometric rather than the energy model. This results in difficulties in their understanding of higher-level molecular biology because they really do not have a solid understanding of how energy works at that level, or how mutational change leads to changes in structure and function. This, in turn, creates difficulties in their understanding of evolutionary processes – the geometric interpretation does not allow for small variation in the binding affinity between molecules or their catalytic activity or substrate specificity. According to the geometric interpretation, two molecules either "fit" or they don't. [We are exploring this in greater detail over the next few months through additional research and there may be an upcoming manuscript available by the time of the CABII meeting so that we can share these more complete results.]

Further Help from the Community

These two broad conceptual areas have implications beyond those explored briefly here. It is our hope our research results will motivate the CAB community to recognize, and development interventions (tutorials) to improve student learning with regards to these foundational concepts, that is, randomness in biological systems and energy as a way of understanding molecular interactions. We anticipate that such research will lead to further CI question development that extend the power of the BCI and other concept inventories, so as to enhance our ability to map students' conceptual understanding in these areas.

The BCI: Results and Disseminiation

The BCI project has resulted in the development of two tools that can be used to map students' conceptual understanding. First, there is the concept inventory itself. The BCI is a multiple-choice formatted instrument that can be administered pre- and post-instruction or as a tool to monoitor student progress at their entry and exit from a program. The BCI does not cover all introductory biology level material at the conceptual level, but rather, it covers several of the really critical "big ideas" that students will need to master in order to truly master and then contribute to the discipline. While its reliabilty is greatest when all items are administered, those teaching can also select only those items they cover in their course at the conceptual level. At least, this is what we are working to develop (a instrument modification tool) on our website. Currently, we are preparing, or have submitted, papers for publication that explore:

1) the statistics related to the BCI's validity and reliability;
2) our research results as they related to what we feel are the two most critical "big ideas" being covered by the BCI as we breifly discussed above

- one paper, currently under review, related to randomness and its impact on the understanding of molecular processes and evolutionary change
- additional research and a paper exploring the geometric vs. energy model of molecular affinity (and related constructs)
- a paper that addresses students understanding of how changes in genes relate to phenotype, and how genes (as abstract objects) behave, according to Mendelian rules

3) a paper exploring issues related to students' use of the rhetoric of science as seen in biology, astronomy, and physics

In addition to these works for publication, we hope to extend the coverage of the BCI in specific directions, one of particular interest is how students view interacting systems, be they pathogen-host, predator-prey, and gene-regulatory factor. Finally, we are working to complete our website so that it is better explains the BCI instrument, the research and development design, and the various applications for it.

In order to analize the vast amount of text-based data collected for this project, we developed a tool that enables us to begin to map students' understanding at the conceptual level. This tool is designed to facilitate content analysis of text-based data. Essentially, we designed a platform independent, web-based tool. Called, Ed's Tools, it can be used in a number of ways and continues to grow in functionality and applications as we continue our work to research, identify and understand students' misconceptions (Garvin-Doxas works not only on the BCI, but in other STEM disciplines in the area of student misconceptions). We designed and developed Ed's Tools in an effort to provide people who specialize in a STEM discipline with an easy means of coding text-based data since researching misconceptions requires a multi-disciplinary team and not all members are familiar with or want to take the time to learn the often complicated social science software commonly used for the analysis of textual data. As we describe in more detail in several of our recent presenations, papers, and on our website (e.g., Garvin-Doxas, et al 2007; Garvin-Klymkowsky & Garvin-Doxas, under review), text data is imported into the program and then coded for patterns in expression that indicate various types of student understanding. This content analytic approach to the exploration of this data can be based on *a prior* categories in a deductive manner or it can be used in an inductive manner where analysts are guided by the questions of:

- What is it that students say in response to these questions?
- What patterns in their responses are present?
- How can (or can) these patterns be interpreted?
- Do these recurring patterns tell us anything about students' conceptual understanding"

Ed's Tools provides a vehicle for the collection of students' natural language in response to essay or short answer questions. Results of analyses using Ed's Tools are being used to collect data related to students' conceptual understanding by professors who simply want a means of breifly "touching base" with their students' non-rote understanding (e.g., a means of discovering how much of what I intend to teach are my students actually absorbing); as a means of researching students' misconceptions; for the purposes of research in natural language processing; etc. in a wide variety of STEM courses at the grade 3-post-secondary levels. (Please see the Garvin-Doxas, Doxas, and Klymkowsky workshop proposed for this meeting as well as our published paper that focuses on the Tool.)

Currently, dissmenation of both the BCI and Ed's Tools consists of our website at www.bioliteracy.net , as well as through presentations and workshops we have been invited to give as well as those made at relevant conferences (e.g., we have spent a great deal of time disseminating both the concept inventory and Ed's Tools at the National Evaluation Association, because evaluators are often in a position to introduce a tool to those working on educational projects). In addition, we have begun to publish additional work related to the research and development of the BCI in relevant journals. We are contacted on a weekly basis with requests from members of the community of biological science teachers for access to the BCI. We request access to their data, but provide access to the instrument to any who request it. We are working this year to conduct a more broad-based pilot test of the final version of the current BCI question set so that we can provide the most accurate statistics possible on its validity and reliability, as well as to ensure that it remains stable across different types of Introductory Biology Course contexts.

Informing and Improving Biology Education

The BCI project was not attached to any classroom intervention and the project goal was to develop a CI that could be used to measure student conceptual understanding at the Introductory level. As a result, the project has focused on designing the BCI in a way that will meet a number of potential uses. An unexpected outcome of our project was the development of Ed's Tools which can be used in a variety of ways for the purpose of mapping students' conceptual understanding. The main use of the BCI (and Ed's Tools) in improving biology education is in the discovery of the misconceptions that arrest student conceptual progress. One can envision, for instance, instructional materials that would teach directly the idea that random processes are taking place all the time, and that the "directed motion" we observe is only a result of selection (whether we select to only "look" at the dye molecules in a cup of water, which then appear to move down the density gradient, or select to follow certain mutations). Similar instructional materials can be envisioned for teaching an energy-based understanding of molecular bonds, which allows for a continuous change in molecular affinity (and therefore molecular evolution).

The primary uses of the BCI are:

- Provide understanding of where students currently are "located" on the conceptual landscape
- Provide a pre- and post-instruction comparative measure of student progress in terms of their conceptual understanding of the "big ideas" addressed by the BCI
- Enable those teaching to make informed decisions about where to dedicate the most class time
- Enable those teaching to make informed decisions about appropriate interventions for students; interventions that specifically address the most commonly-held student misconceptions
- Provide accurate information about the success of a particular teaching intervention with regard to a particular "big idea" or a particular subset (e.g., the diffusion component of student understanding of random) at both the local level and across institutions
- Provide reliable data that can be employed in course evaluation; evaluation across similar courses; track changes in student understanding based on their tenure in the program (programmatic evaluation)
- Provide insight into the way specific teaching interventions successfully address particular conceptual-level understanding

For the range of uses of Ed's Tools beyond what is discussed in the previous section, please see our paper focusing specifically on this tool (Garvin-Doxas, Doxas, and Klymkowsky, 2007).

Community Assistance

While it is not in the scope of the BCI project to conduct research on the ways in which the instrument(s) contribute to improved biology education, proper use of CI's has been shown to enhance STEM education (as has been well-documented with the Force Concept Inventory (FCI) over a more than 20 year period in Physics Education Research). Ultimately, we hope the BCI becomes widely adopted and shown to contribute to improved educational approaches in the biological sciences. Continued communication within the broader community is the only way this can occur. For example, Mike Klymkowsky recently was invited to talk about results from the BCI at the annual International Meeting for the Systems Biology (held in Long Beach, CA). Not only did this bring an important message to group of biologists, but also it revealed areas of research that an expanded BCI should address. Certainly the bulk of research demonstrating the influence of the FCI has been conducted by those other than its developers. This is also true of its wide-spread adoption and dissemination: educators other than Halloun and Hestenes (1995; Halloun, et al, 1992) contributed to its adoption (e.g., Mazur, 19??; Hake, 1989). If we follow the model of the Force Concept Inventory, we can expect that dissemination of results from the BCI will encourage science educators to examine the conceptual understanding of the students in the areas currently "covered" and others, and to provide a model for biology education researchers to build alternative CIs that address their own specific needs, e.g the Force and Motion Concept Exam (FMCE) and the Brief Electricity and Magnetism Assessment (BEMA).

Challenges

We have previously discussed the challenge to CI development posed by students' internalization of the rhetoric of science and are currently working on a paper exploring this issue in further detail. Students, particularly those in the biologial sciences, are so emerssed in the traditional, rote-memory approach to science learning that it is especially challenging to create essay, interview, and CI questions that access their conceptual level understanding and the mental models that drive it (this is discussed in the Garvin-Doxas, et al, 2007 paper that is in press; see also the Garvin-Doxas and Klymkowsky, 2007 paper prepared for the initial CAB meeting at www.bioliteracy.net).

From a more pragmatic perspective, perhaps our greatest challenge comes from the resemblance that CIs bear to tests. Our focus on the identification of verbal markers (generally descriptions or phrases found in students' natural language) has contributed to misconceptions about CI's. With the recent emphasis on accountability and testing, STEM education research has come to focus on aspects of test theory that are inappropriate (or have no bearing on) the creation of a diagnostic instrument. As a result, many recent efforts to develop CIs show a marked tendency to create valid tests rather than a diagnostic instrument. This tendency to apply the psychmetrics essential to test development to the construction of CI's is further contributed to by the typical multiple-choice format used by both. A CI is designed to diagnoise student misconceptions; to map where they are in terms of the conceptual landscape. Not to test students' skill and knowledge. This is not to say that well-constructed tests have no place in STEM education. Excellent tests are essential to the improvement of undergraduate education. At the same time, tests measure different things from CI's and it is important to keep their purposes in mind when determining the appropriate instrument to be used in any given situation.

Tests are basically designed to answer the question "what percentage of the desired knowledge and skills in this field has this student acquired?". CIs are meant to answer the question "what is the probability that this particular student uses that particular conceptual construct when solving problems in this field?". These same questions can also be asked from the point of view of an ensemble of students (rather than the individual student). From that point of view tests are meant to rank the students in the ensemble by the their skill and knowledge, while CIs are meant to report the percentage of students in the ensemble that use a particular conceptual construct. Tests are therefore 1) uni-dimensional (a test may probe different dimensions, but the final grade is of course uni-dimensional) 2) monotonic (A > B > C > D) and, as much as possible 3) linear (A-B = B-C = C-D). To ensure these properties, test developers look at statistical measures like discrimination (ie. how close can two scores be before we can no longer assure that the higher score indeed represents higher performance) and item difficulty (so that items of different difficulty on a given subject are included in the test, to insure that answering twice as many questions really requires twice the level of performance). Additional measures of reliability look at the correlation between individual items and the instrument as a whole as an additional check on the appropriate level of difficulty of each question. Validity is usually tested by making sure that grades in (a large number of) relevant courses are distributed over a wide range and centered in the statistical mean (which is a perfectly reasonable criterion for a test; having the grades over a large number of different classes all clustered between 35% and 45% is a sure sign that the test does not explore the full variability of student performance).

Necessary as these statistical properties are for tests, they are mostly irrelevant (and sometimes even counterproductive) for Concept Inventories. CIs are by nature multidimensional since what we really want to know is the misconceptions that a student holds, not some average of over all misconceptions (what we really want to know is what specific instructional material to assign to a student in order to address his/her misconceptions; a measure of the student's average performance level is not at all informative on that task). Since they are multidimensional, monotonicity and linearity are largely not applicable, and in any case the percentage of students that answers a question correctly is not an appropriate weighting factor (the vast majority of the students can, and often do, harbour the same misconception even after repeated instruction; this is the very essence of misconceptions). Finally validity has to do with the rate of misses (how many students hold a particular misconception and we failed to identify them) and false positives (how many students we identified as holding a particular misconception but they don't). These measures are best arrived at by conducting large numbers of interviews and comparing the statistics from the interviews with the statistics of the instrument, and adjusting the instrument until the statistics match, which is a long and expensive process.

Finally, we have recently begun to explore the best ways to describe the meta-structure that scaffolds the various distracters to the concept questions we have developed for the BCI because we believe that it will enhance our ability to communicate the nature of student understanding both within the CAB community and with the broader community of biology educators. Unlike typical test questions, CI distracters are based on whatever misconceptions students commonly hold and which have been identified through research into student understanding (rather than on expert-driven choices). Essentially, we are working to understand whether or not there is a pattern among CI distracters that characterizes the nature of the conceptual landscape we seek to map. In most cases, students' commonly-held misconceptions do not represent a linear progression in terms of level of abstraction in their understanding. In other words, the distracters found in CI questions do not form a linear progression from e.g., the knowledge level of understanding or abstraction to the evaluation level described by Bloom's Taxonomy (Bloom, et al, 1959). In an effort to understand how we *can* categorize the distracters found in the BCI, we have become interested in how we might better describe and explain mapping student conceptual understanding as it is represented by distracters developed for CI's, by using something like the Progressive Learning Model (Michail, et al, 2006). We feel that the ability to articulate an organizational schemae that represents the conceptual landscape will assist us in our efforts to address other challenges that CI developers face as well.

We are particularly interested in exploring more about the nature of "big ideas" in biology education; the psychometric models most appropriate and relevant to CI questions; and how to describe the conceptual landscape being mapped through the use of BCI and other similar instruments.

References:

Anderson DL, Fisher KM, Norman GJ (2002) Development and evaluation of the conceptual inventory natural selection. *Journal of Research In Science Teaching 39: 952-978.*

Bloom, B., Englehart, M., Furst, E., Hill, W., and Krathwohl, D. (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals, by a committee of college and university examiners. *Handbook I: Cognitive Domain.* New York; Longmans, Green.

Garvin-Doxas, K., Klymkowsky, M. W., and Elrod, S. (2007; in press). 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. *Life Science Education.*

Hake RR (1998) Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *Am J Physics 66: 64-74.*

Hestenes D, Halloun I (1995) Interpreting the FCI. *The Physics Teacher 33: 502-506.*

Hestenes D, Wells M, Swackhamer G (1992) Force concept inventory. *The Physics Teacher 30: 141-166.*

Klymkowsky, M.W. (2007; in press). Teaching without a textbook: a strategy to focus learning. *Life Science Education*.

Lynch, M. (2007). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. USA 104, suppl 1: 8597-604*.

Mazur, R. (1997). *Peer Instruction.* NJ; Prentice Hall.

Michail, N., Teal, G., Basta, J. (2006). Progressive Learning Process Model – Interpretive methodolgical framework for human systems inquiries. Proceedings of the 50th Annual Meeting of the International Society for the Systems Sciences (ISSS), Sonoma, CA.