Diagnostic Question Clusters to Improve Student Reasoning and Understanding in General Biology Courses: Faculty Development Component

CAB II Meeting, January 2008

Charlene D'Avanzo[1], Deb Morris[2], Andy Anderson[3], Alan Griffith[4], Kathy Williams[5], and Nancy Stamp[6]. 1. Hampshire College, 2. Jacksonville Community College, 3. Michigan State University, 4. Mary Washington University, 5. San Diego State University, 6. Binghamton University, SUNY, Binghamton, NY

Introduction

Since the last Conceptual Assessment in Biological Sciences meeting, we have received CCLI support for a program that targets two challenges to the teaching of General Biology: 1) Most students do not address biological questions with the principles and reasoning used by biologists and 2) Most faculty do not teach students how to use the principles and thinking of practicing biologists. To address these challenges, our proposal centered on a set of interrelated biological *Diagnostic Question Clusters* (*DQCs;* e.g. Wilson et al. 2006) designed to "hook" biology faculty to question and learn about their students' understanding of core biological concepts and ways of thinking about biology (see example below). As Andy Anderson, Joyce Parker, and John Merrill explained in CABS I (Anderson et al. 2007), one set of the *DQCs* they have developed concern tracing energy and matter through three levels of biological complexity (subcellular-organismal). In this new project we will expand this framework to a fourth level, ecosystem, and focus on this in our workshops for faculty.

Although the *DCQs* developed by Anderson et al. (2007) are based on extensive study and research, even the best diagnostic tools will not on their own lead to better biology instruction. College faculty need a great deal of help – using such tools, interpreting findings from them, and incorporating effective practices that help their students overcome common misconceptions and poor biological thinking. Therefore, in this project we are linking a set of well-researched diagnostic tools for teaching biology with a faculty development program. Specifically, we will study faculty use of *DQCs* developed to evaluate students' understanding of core biological ideas and ways of thinking. Faculty who teach General Biology in diverse institutions will use these *Question Clusters* as a means to improve their teaching via a faculty development program based on our experience with *TIEE* faculty development (see below) In this way, we will be able to study how faculty use such diagnostic tools in a range of classroom and institutional settings. Results from this study will help us understand whether diagnostics like the *Cluster Questions* are potentially effective tools for improving introductory biology teaching and learning - and the associated types of faculty development programs that we need.

Table 1. A question from our cluster on cellular respiration. Each foil represents a common incorrect response to open-ended questions and can be seen as indication of the problem as indicated in brackets**.** For example, when asked to trace matter, students commonly give an answer about energy. The correct answer is A, in bold. For more information about sets of *DQC*s, contact Andy Anderson (andya@msu.edu).

Jared, the Subway man, lost a lot of weight eating a low calorie diet. Where did all the fat / mass go? **A) The mass was released as CO_2 and H_2O.** B) The mass was converted to energy and used up. [*mass-energy conversion*] C) The mass was converted to ATP molecules. [*tracing matter]* D) The mass was broken down to amino acids and eliminated from the body. [*tracing matter*] E) The mass was converted to urine and feces and eliminated from the body. [*failure to move to cellular level*

Faculty resistance and reluctance to changing their teaching

Real change in basic biology teaching clearly necessitates faculty understanding how and why students so poorly understand important biological ideas and concepts. However, most faculty have limited formal knowledge about teaching and learning, and therefore engaging them in topics such as misconceptions and biological reasoning is a big step. How do we deal with this dilemma? What will entice faculty to take the time to become educated about problems so fundamental to teaching biology and ecology? To address this, we look to other faculty development programs that focus on science teaching and learning at the undergraduate level.

TIEE: Eight years ago, D'Avanzo et al. (2007) established *Teaching Issues and Experiments in Ecology (TIEE*), a peer reviewed publication of the Ecological Society of America (ESA) designed to help faculty include more student-active approaches in the courses, including lecture. A Research Practitioner program for faculty using *TIEE* is the basis for the new *DQC* project.

The *TIEE* Research Practitioners Project uses the concept of "action research," a mode of inquiry from education and related social sciences. Action research is done by practitioners (e.g. teachers) who ask questions and design studies that will directly inform their practice (Stenhouse 1975). It is different from evaluation in part because practitioners work in teams on common issues. Though done infrequently in higher education, college faculty have effectively used methods of action research to improve their own teaching even without grounding in the critical theory foundation of "pure" action research (Kember and McKay 1996; Webb 1996). Similarly, in *How People Learn,* Bransford et al. (1999) specifically recommend that teachers conduct research via formative assessment. Throughout that volume the importance of making students' thinking visible through frequent assessment, feedback, and revision, is emphasized, although the authors also acknowledges that " the knowledge base on how to do this effectively is still weak". Notably, these ideas are not new. John Dewey

75 years ago complained about the separation of education research and the disciplines being studied, saying that good teaching depends on "direct participation of those directly involved in the research" (Dewey 1929,47).

Interest and discussion of research on teaching conducted by college faculty themselves has greatly increased in recent years (e.g. NRC 2002, Fox and Hackerman 2003). Handelsman et al. (2004) use the term *scientific teaching* to describe teaching based on theories of teaching and learning and informed by empirical evidence. *Classroom assessment* (Angelo and Cross 1990) and *classroom research* (Cross and Steadman 1996) are increasingly popular ways for college faculty to engage in ongoing, practical inquiry into student learning. Clearly the terminology – evaluation, assessment, research, etc. - is confusing. We used the term "practitioner research" to describe inquiry conducted by faculty on their own practice, encompassing action research, classroom research, and scientific teaching.

In a 2005 summer *TIEE* Research Practitioner workshop, faculty formed self-selected groups based on common issues they wished to investigate in biology and ecology courses (D'Avanzo and Morris 2007). For instance, one team of four faculty from a range of institutions used pre- and post tests and other measures to study their students' progress working with data. In the two semesters after the workshop, faculty teams continued to work together via conference calls. During that academic year, these faculty carried out studies they designed in consultation with each other, collected and analyzed various types of data, pondered the meanings of their findings, and reworked their experimental designs. At the 2006 annual meeting of the Ecological Society of America, they presented six posters (several coauthored) describing their research and how the process of doing practitioner research was changing them as teachers. This led to eight publications in Vol. 5 of *TIEE* (refs). We consider this excellent progress.

In regard to faculty development, our major findings from this project are:

1. Importance of teams: Especially relevant to the new project is the finding that the team aspect of the RP program was essential. Faculty said that they likely would not have continued to ‘stick it out’ without the support and help of their team members. Several emphasized that, even though they had not participated in the conference calls and listserv discussions as often as they would have liked, just knowing that others were out there working on the same issues was motivating, plus they knew they could get help and support if needed.

2. Link between student-active approaches and aspects of scientific thinking: Also important is the deliberate link between use of a particular active approach (e.g. repeated group work with tables and figures) and gains in identified aspects of scientific reasoning (generalization from data in figures and tables to particular concepts or information). Faculty connected gains they measured in student learning with such focus and repetition.

3. Limited scope of study and iteration: A related point is that faculty limited their research to specific aspects of their teaching. They asked quite specific questions, developed ways to address these, examined their findings along the way, and used this feedback to modify their teaching over the semester. This is the essence of formative evaluation. However, our experience with *TIEE* is that most faculty do not engage in this practice, even when we specifically and repeatedly urged them to use the *TIEE* modules (D’Avanzo et al. 2005).

Lessons Learned From Other Science Faculty Development Programs

We are in the first stages of analyzing the literature on science faculty development programs and barriers to change in teaching for faculty (interestingly, these studies appear to be limited). From several relevant papers (Sunal 2001, Akerlind 2007, Dancy and Henderson 2007) key findings are:

Table 1. Key aspects of workshops and associated efforts to help college science faculty improve their teaching.

1. Interactions with other faculty are critical – faculty need to work with other teachers with similar teaching situations and whom they can trust; periodic interactions need to be scheduled ahead of time.
2. Action research is an important element for many faculty.
3. Programs should introduce faculty to education research.
4. Change begins with the goal to be accomplished, not the barriers.
5. Incremental change is the norm and a reasonable expectation.
6. Sophisticated teaching development means continually increasing one's understanding of what works and doesn't work for students.
7. Since teaching situations are very different, we need to provide easily modifiable materials.
8. Collegial and administrative support at the home institution is important.
9. Faculty should be seen as partners working with us on this program
10. We should acknowledge that change is difficult and support faculty, not criticize them.

Implications for our Biology Concepts Faculty Development Program

The specifics about how we will suggest that faculty use the *DQC*s in their biology course will take shape as we plan the 2008 summer workshop. So that we have common measures across institutions, we will likely ask that students be given a small set of diagnostic questions as pre and post tests in association with the ecology component of the course. Faculty will then decide now to use the rest of the DQCs and associated student-active (think-pair-share, concept mapping, jigsaw etc.) and formative evaluation approaches. The workshop will be designed to participants make these decisions to minimize on-the-fly decisions during the busy semester.

The findings outlined in Table 1 confirm our commitment to these aspects of our own program to help faculty use *DQC*s to improve their teaching of introductory biology:

- Faculty will work in teams most likely by institutional setting. Therefore workshop participants should include several faculty from R-1 universities, other universities, four year colleges, and community colleges. In addition, teams must continue to work together (e.g. via conference calls) over the academic year.
- Faculty should be introduced to numerous effective ways to do educational research in their courses (e.g. pre/post tests, surveys, extended response questions with associated rubrics).
- The Research Practitioner model (faculty doing action research on their teaching) is a good one. A related point is that faculty need advice developing strong education research designs.
- A critical component of the Research Practitioner model is scholarly publication of findings (posters, talks, papers). They will need help finding such venues.
- At the start faculty should focus their study on a limited aspect of the course (in our case on students' understanding of carbon and energy dynamics).
- Faculty need a range of options for using the materials (e.g. student-active approaches for large-small classes).
- In addition to the diagnostic tests themselves, and the associated framework, faculty need a variety of active teaching approaches specifically linked to the diagnostic material (e.g. how to use a multiple-choice 'clicker' question several ways in different settings)
- Follow-up workshops are critical

However, some of these recommendations seem at odds with the goals of national biology concepts inventories. For example, a very compelling component of the Force Concept Inventory studies is the large number of students taking the identical FCI test (Hake 1998). However, how do we encourage faculty to modify diagnostics for use in their courses and yet ask them to give students identical tests in the same way? We will continue to struggle with this question.

References

CW Anderson, J Merrill, and J Parker. 2007. Teaching and learning biology at the undergraduate level conceptual assessment in biology group. CABS I March 2007 Boulder CO, meeting

Angelo TA and Cross KP. 1993. Classroom assessment techniques: A handbook for college teachers (2nd ed.). San Francisco, CA: Jossey-Bass.

GS Akerlind. 2007. Constraints on academics' potential for developing as a teacher. Studies in Higher Education 32: 21-37

JD Bransford, Brown AL, and Cocking RR. 1999. How people learn: brain, mind, experience, and school. Washington, DC: National Academy Press.

C. Callow-Heusser, Chapman HJ, and Torres RT. 2005. Evidence: An essential tool. Arlington, VA: National Science Foundation.

KP Cross and Steadman MH. 1996. Classroom research: Implementing the scholarship of teaching. San Francisco, CA: Jossey-Bass.

C. D'Avanzo & D. Morris. 2007. Investigating your own teaching: faculty as research practitioners. Academe, accepted for publication

C D'Avanzo et al. 2006. Design and evaluation of TIEE: A peer-reviewed electronic education resource. Frontiers for Ecology and the Environment 4: 189-195.

C D'Avanzo. 2003. Application of research on learning to college teaching: ecological examples. BioScience 53: 1121-1128

C. D'Avanzo. 2003. Research on learning: Potential for improving college science teaching. Frontiers for Ecology and the Environment 1: 533-540

J Dewey. 1929. The Quest for Certainty: A Study of the Relation of Knowledge and Action (New York: Minton, Balch, 192

D Kember and McKay J. 1996. Action research into the quality of student learning: A paradigm for faculty development. Journal of Higher Education 67: 528-554.

J Handelsman et al. 2004. Scientific teaching. Science 304: 521-522.

RR Hake. 1998. Interactive-engagement versus traditional methods: Asix-thousand student survey of mechanics test data evaluation of active learning laboratory and lecture curricula for introductory physics courses. American Journal of Physics 66: 64–74.

C Henderson and M Darcey. Physics faculty and educational researchers: divergent expectations as barriers to the diffusion of innovations. American Journal of Physics (Physics Education Research Section). In Press.

J McNiff. 2002. *Action research: Principles and practice*. London: Routledge Falmer.

National Research Council. 2002. Scientific research in education. Washington, DC: National Academy Press.

D. Schon. 1995. The new scholarship requires a new epistemology. Change 27: 26-34.

L. Stenhouse. 1975. Introduction to curriculum research and development. London: Heinemann

G Webb. 1996. Becoming critical of action research for development. In: Zuber-Skerritt O. (Ed.). New directions in action research. London: Routledge Falmer.

D Wilson, CW Anderson, M Heidemann, JEMerrill, BW Merritt, G Richmond, DF Sibley, and JM Parker. Assessing Students' Ability To Trace Matter In Dynamic Systems In Cell Biology. Cell Biol Educ 2006 5: 323-331.

O Zuber-Skerrit. 1992. Professional development in higher education: A theoretical framework for action research. London: Kogan Page.

DW Sunal. 2001. Teaching science in higher education: faculty professional development and barriers to change. School Science and Mathematics 101: 246-257

Genetics Concepts Inventory (GenCI) Development

Susan Elrod, Ph.D.
October 15, 2007
Biological Sciences Department, California Polytechnic State University,
San Luis Obispo 93407 (selrod@calpoly.edu) 805/756-2875

Abstract

A genetics concept inventory (GenCI) is under development as a tool to measure students understanding in different learning environments. The motivation for developing the GenCI is as one of several assessment strategies for evaluating an innovative literature-based case study (LBCS) approach for teaching introductory college-level genetics at California Polytechnic State University, San Luis Obispo (Cal Poly). The initial development of the GenCI and the LBCS approach are described in a paper presented at the Conceptual Assessment in Biology workshop (1). The LBCS approach also incorporates a significant scientific information literacy component (2). This paper outlines the main concepts (i.e., big ideas) in genetics, the further development and use of the GenCI assessment tool and the challenges associated with its development and use.

Big Ideas in Genetics

The big ideas in genetics that were used to develop some of the GenCI questions are summarized in Table 1. There are four main categories: 1) nature of the genetic material, 2) gene expression and regulation, 3) transmission, and 4) variation and evolution. Current GenCI questions are focused primarily in areas 1 and 3, and additional questions in areas 2 & 4 are in development.

Table 1. Major Concepts in Genetics[1]

Major Concept Area	Information Content (Structure)	Information Flow (Process)
1.0 Nature of Genetic Material	Genes are the basic unit of biological inheritance Biological inheritance information is located in the chromosomes, which	Genes contain the information for products, such as proteins (polypeptides) or RNA molecules, that contribute to the determination of traits within cells and

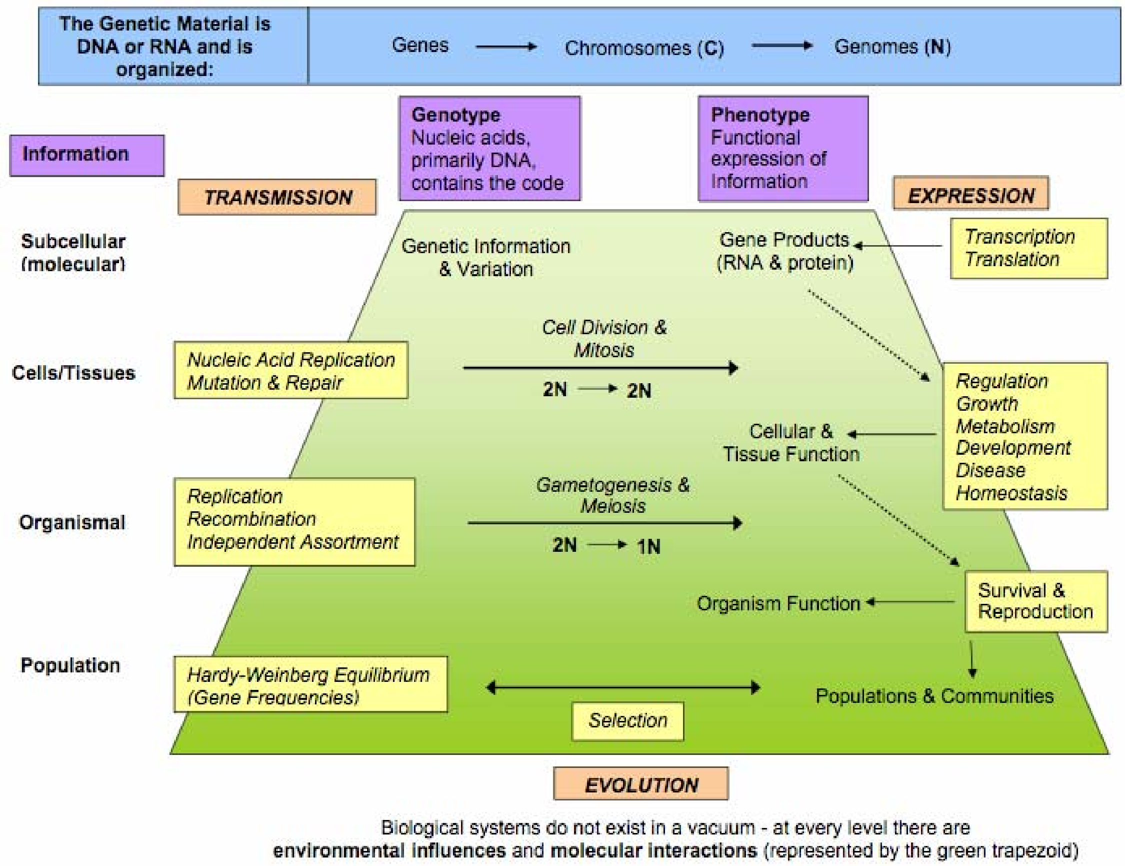
	are made of DNA and proteins and located in nucleus of eukaryotic cells; all cells have chromosomes (all living organisms are made up of cells) DNA is composed of deoxyribonucleotides (ACGT); RNA of ribonucleotides (ACGU); nucleotides are composed of a 5-carbon sugar, base, phosphate group; proteins are comprised of amino acid subunits Chromosomes in diploid organisms are grouped in pairs, which called homologous chromosomes; homologous chromosomes contain gene pairs, located in the same position on each homologue (locus); mitochondria and chloroplasts contain their own DNA There are two types of chromosomes, autosomal and sex (e.g., X and Y); all animal cells contain sex chromosomes; eggs of mammals contain a single X and sperm can contain either an X or a Y	organisms; enzymes are an example of one class of proteins; enzymes catalyze chemical reactions within cells; chemical reactions result in the production of molecules used for various cellular processes (e.g., amino acids used to make proteins); abnormal gene expression can lead to disease (e.g., cancer); one gene, one gene product (not enzyme, protein, polypeptide) Most traits are determined by the action of many different genes, as well as the interaction of gene products with their environment; genes and the environment interact to determine phenotype
2.0 Gene Expression and Regulation	The information within DNA molecules contains the code used by cells to produce RNA and/ or protein molecules The genetic code refers to a triplet sequence of DNA that codes for a specific amino acid during protein synthesis; one codon specifies one amino acid in a polypeptide chain The physical composition of an organism is called the phenotype; genetic composition called the genotype	The central dogma describes how genetic information functions within the cells of an organism; transcription results in the production of RNA; translation results in the production of proteins Inside cells, different genes are expressed over the course of development, in different cell types and in response to environmental stimuli specifying the cellular type, shape and function
3.0 Transmission	Somatic cells of an organism carry the same number of chromosomes and the same inheritance information Gametes carry half the chromosomes and, consequently half the inheritance information; the combination of alleles (or inheritance information) in each gamete is unique and different from the parents	Meiosis is the process by which genetic information is passed from generation to generation in sexually reproducing organisms; however, mitosis (asexual reproduction) is used by some organisms as a means of reproduction Each homologous chromosome is inherited from one of the diploid organism's parents'; homologous chromosomes separate from one another during meiosis (Principle of Segregation); different chromosomes (and thus genes and alleles) will separate independently of one another during meiosis (Principle of Independent Assortment), however, genes (alleles) located close to one another on a chromosome are linked and more are likely to be inherited together

		than those farther apart or on another chromosome; recombination disrupts linkage relationships A zygote is formed by the process of fertilization of two gametes in both plants and animals Plants also use mitosis and meiosis for production of daughter cells and gametes, respectively; bacteria reproduce by a process known as fission which is different from mitosis as well as meiosis
4.0 Variation and Evolution	Mutations are heritable changes in the sequence of DNA; can involve single or a few DNA base pairs or large regions of chromosomes; spontaneous or environmental agents (chemicals, radiation) Genes may take on different forms, which are called alleles; alleles are composed of slightly different sequences of DNA and encode gene products of different sequence as determined by the changes in DNA; gene products from different alleles may have drastically or moderately different function For a given gene, diploid organisms can be homozygous (same alleles) or heterozygous (different alleles) Diploid organisms may contain only two different alleles for any given gene at a specified locus Genes and environment together are responsible for the phenotypic variation between species	Processes that cause individuals of a species to be genetically unique: recombination, independent assortment, and mutation Allelic relationships can be complete dominance and recessive, incomplete dominance, semi-dominant, co-dominant; dominance means that trait is observed in the phenotype of a diploid organism regardless of the nature of the homologous allele; recessive means that trait is not observed in a diploid organism if the homologous allele is dominant Cells contain mechanisms for repairing damage to DNA Natural selection acts upon genetic variation within populations resulting in changes in gene frequencies, and thus evolution

[1] Major conceptual areas were identified and the information content and flow were elaborated. Concept descriptions were modified from 3- 11 based on a framework developed by the Michigan State University group (12).

Hott *et al.*, (2) have described six genetics conceptual areas in the context of an introductory biology course for non-science majors: 1) the nature of the genetic material, 2) transmission, 3) gene expression, 4) gene regulation, 5) evolution and 6) genetics and society. A visual representation of genetics concepts was developed using a variety sources to enable better understanding of the relationships between major concepts (Figure 1).

Figure 1. Genetic Concepts



This figure was developed in discussion with Joyce Parker, John Merrill, Brett Merritt and Merle Heideman (Michigan State University) and Mike Klymkowsky and Kathy Garvin-Doxas (University of Colorado, Boulder).

The Genetics Concept Inventory (GenCI): Results and Use

The GenCI was given in its original form as a pre- and post-test to upper division genetics students during Spring 2006 (Adv I). A revised version was given as a post-test in a lower division cell and molecular biology course during Fall 2006 (Intro) and as a pre-test in an upper division genetics course during Fall 2007 (Adv II). It will be given again as a post-test in the same Fall 2007 genetics course. Overall results are shown in Table 2. The Adv II students had the entire first week of class to take the test, which may explain why their average score was slightly higher. A recommendation for future pre-tests is to have students take it before the class period.

Table 2. GenCI statistics

Test	Average	Standard Deviation	Range (High/Low)
Pre-test Adv I (n=49)	50% (38/76 points)	11	64/16
Post-test Adv I (n=32)	65% (49/76 points)	9	68/25
Intro Post-test (n=107)	63% (25/40 points)	5.5	36/10
Pre-test Adv II (n=50)	59% (23/39 points)	5.7	35/14

Adv I = Version 1.0 given in Genetics course during Spring 2006 contained 38 questions each worth 2 points; Intro and Adv II = Version 2.1 given in Intro cell and molecular course in Fall 2006 contained 40 questions each worth 1 point.and in Genetics course during Fall 2007 contained 39 questions each worth 1 point.

Analysis of Selected GenCI Questions

A few GenCI questions are highlighted in this analysis, results of which are shown in ascending order of version 1.0 in Table 3. The question numbers changed between version 1.0 and 2.1, and some questions changed within version 2.1 as indicated in the data tables and discussion.

Table 3. GenCI questions of interest

Q # (A)	***Q # (B)***	***Question Content***	***Intro post %***	***Adv II Pre %***	***Adv I Pre %***	***Adv I Post %***
16	8	Nature of allele	37	46	31	81
26	29	Relationship of genetic information in egg and ovary cell after oogenesis	41	44	49	50
29	12*	Cell types that contain genetic information for eye color	73	34	20	52
31	19	Nature of genetic recombination	22	18	42	75
34	34	Products of translation are proteins	31	52	37	75
35	33**	Products of transcription	77	18	59	88

*A= Adv I GenCI (version 1.0); B = Intro & Adv II (version 2.1)****; * = question revised from 1.0 to 2.1; ** = question revised from Intro to Adv II***

Question 16/8 asks students to identify the correct response to the question, “what is an allele?” Less than half the Intro, Adv II pre and Adv I pre students choose “an alternate form of a gene.” Other answers chosen are “part of a gene,” “one of two genes in a diploid organism,” or “a region of chromosome where a gene is located” (Table 3). It may be easy to see why they would choose “one of two genes in a diploid organism” but interviews will help us better understand this result. A corollary question that asks students what the maximum number of alleles of a single gene that can exist within a diploid organism with 6 chromosomes reveals that students may not fully understand the meaning of the word allele: a minority 20% of Adv II students chose the correct answer (2), while the majority of students chose 12 (58%).

There are several questions that deal with the results of chromosome and genetic information consequences of mitosis and meiosis. Students show a high degree of understanding regarding the chromosome and genetic information consequences of mitosis; that is, on the pre-test 80% correctly state that the chromosome number is the same before and after mitosis and 86% correctly state that the genetic information between daughter cells is the same after mitosis. And, students are very clear on the fact that egg and sperm cells from the same organism have the same number of chromosomes (data not shown). However, students are essentially split on their responses regarding the genetic information content of ovary and egg cells after meiosis (Question 26/29, Table 3). Perhaps students are confused with the intent of the question, i.e., an egg cell may be interpreted as the same as an ovary cell, or perhaps they truly are confused about the relationship between the genetic information in these two different cell types. Interestingly, greater than 65-70% of students accurately answered that an egg cell has half the chromosomes as an ovary cell after meiosis. Interviews with students will shed more light on the reasons for these apparently incongruous answers. Also, more robust statistical analysis (beyond that available in the Blackboard learning management system) will allow correlations between pre- and post-test answers. For example, it will be informative to know whether the students who still think that egg cells contain the same genetic information as ovary cells are the same students who cannot correctly identify the number of chromosomes in an egg cell.

A breakdown of specific responses to questions 29/12 is shown in Table 4. This question relates to whether or not students understand that all cell types contain the same genetic information, a misconception identified from the 9th grade/high school research. It appears that a significant number of college juniors maintain this idea even by the end of the term. While 55% of Adv I students start the quarter thinking that only gametes contain the genetic information for eye color, a large number of students (32%) still think this at the end of the term. In the Adv II pre-test results, 30% of students think gametes are the only cells than contain eye color genes; however, 16% think it is eye and gamete cells and 34% give the correct answer (all cell types listed). Interestingly, responses to a corollary question that asks whether genes are found in all cell types in an organism show that 78-80% of Adv I and Adv II students answer correctly. Student interviews as well as statistical correlations will be required to further analyze this discrepancy. In this case, it would be informative to find out whether students making up the 32% who still think that gametes are the only cell type to contain eye color genes are the same students who are confused regarding which cell types contain genes.

Table 4. Responses to Question 29 and 12

Possible Answer	Adv I Pre %	Adv I Post %	Adv II Pre %
All	20	52	34
Eye & gamete	12	0	16
Gamete only	55	32	30

With respect to question 31/19 regarding recombination (Table 3), students usually choose “the process of crossing over during meiosis” over the correct response of “sorting of alleles into new combinations.” This is perhaps understandable as these two concepts are often discussed at the same time and the word recombination and crossing over may also be used interchangeably.

The next set of questions deal with the products and location of translation and transcription (Question 34 and 35/33 in Table 3, respectively). Intro results indicate that 31% of students correctly identify proteins as the products of translation, and 37% or 52% of students (Adv I and Adv II, respectively) do the same (Table 5).

Table 5. Responses to Question 34 (products of translation)

Possible Answer	**Intro Post %**	***Adv II Pre %***	**Adv Pre %**	**Adv Post %**
Proteins	31	52	37	75
Messenger RNAs or ribosomal RNAs	17	16	33	9
Amino acids	13	22	18	0
DNA	16	10	8	3
Amino acids and proteins	22	nd	nd	nd

Fisher (1985) observed that at least half of students in a similar study responded with "amino acids" to a similarly-worded question. That isn't the case here: less than or close to 20% give this response. And, there also appears to be confusion between mRNAs and proteins suggesting that students may be simply mixing up the terms translation and transcription. Fisher's question was worded slightly differently and used "activating enzymes" instead of "proteins" which may also be a factor. Alternatively, perhaps instruction has changed over the past decades to clear up the misconception originally described. It may also be that students really don't know what happens in translation. Student interviews on this topic, including the origin of amino acids will be instructive. With respect to transcription (Question 35/33 in Table 3), Intro and Adv I students have a generally good grasp that mRNA molecules are the proper product. In the Adv II scenario, rRNA was added as a possible choice in a multiple answer question and, in this case, few students (only 18%) identified both mRNA and rRNA as products of transcription.

In the Adv II version of the exam, students were asked to choose where within cells mRNAs, proteins and amino acids were synthesized in three separate questions. Around 60% of students know the correct compartment for mRNA and protein synthesis but the class is split on the location of amino acids synthesis (Table 6). Given that students generally take genetics before biochemistry, their understanding of metabolic pathways may be limited, including those involved in amino acid production. Finally, perhaps adding a question that asks students to identify which cellular process gives rise to amino acids would also be useful.

Table 6. Adv II responses to where mRNAs, proteins and amino acids are made.

Possible Answer	mRNAs	Proteins	Amino Acids
Ribosomes	32	62	46
Free in the cytoplasm	8	30	40
Nucleus	60	6	14

One of the questions asked students to complete the phrase, "One gene encodes…" in a short answer format in Adv I pre and post test, then those short answers were used to construct multiple choice options for the Intro and Adv II test takers. The answers students gave with the corresponding percentages are summarized in Table 7. Common Intro and initial Adv 1 or II answers are that genes encode traits or a single protein. The responses to this question at the Adv I post-test reveal that student thinking becomes more complex over the course of the term: after class, they were less likely to say that a gene encoded a trait and more likely to say that a gene encoded many proteins or a transcriptional unit that could produce many different products. However, it is disturbing that the percentage of students who still believe that one gene encodes one protein or polypeptide increased over the course of the quarter. Perhaps they are still holding on to the famous phrase "one gene, one enzyme." Interestingly, three responses to this short answer Adv I post-test stated that a gene encodes many amino acids. This ties back to Question 34 regarding understanding of translation. Even though there were no students giving amino acids as the products of translation in question 34 at the end of the quarter (Table 5), a few revealed a possible confusion or lack of full understanding in this short answer question.

Table 7. Responses to question, "One gene encodes…"

Possible Answer	Intro Post %	*Adv II Pre %*	Adv I Pre %	Adv I Post %
A trait or characteristic	35	32	43	10
Many traits	14	14	10	6
One enzyme	-	24	0	10
One protein or polypeptide	27	-	19	30
Many proteins	-	6	0	23
A transcriptional unit…	20	24	0	20
Other (short answer only)	-	-	28	1

The GenCI questions for all versions of the test will be provided at the CAB II meeting. Trials of the GenCI have been run at Michigan State University and data analysis of those results will proceed in the next 6 months. Additionally, further question development is underway. Approximately seven questions have been removed because students in all trials gave greater than 75% correct response or it was deemed that there were other more important conceptual areas to include. A plan for student interviews is being developed to better understand some of the results described above. After interviews, we plan to create another modification of the test with modified as well as new questions to pilot.

Challenges

The major challenges discussed at the CAB I meeting in Boulder, CO are still relevant. These challenges include creating clear and unambiguous test items, and writing test items that measure conceptual understanding not just structural or definitional understanding. Technological challenges include statistical analysis (validity and reliability testing; question pairing and correlations) and online test administration that is available to students at a variety of campuses that allows for access not only to the exam, but also to the raw data. It is also challenging to interpret the data, especially if multiple answer questions are used. Multiple answer questions are useful because they provide flexibility to allow students to choose a variety of answers according to what they are thinking. However, it is difficult to know whether or not they realized the question may have multiple answers. A worthy discussion surrounding the use of multiple answer questions would benefit this work. It is also difficult to know whether results from a particular test item indicate a serious misconception or simply a misunderstanding. Certainly interviews help in these cases, however that cannot be done for each test taker. A clearer definition of misconception in biology will be helpful in the development and use of these types of inventories. Additionally, the distinction between misconception and misunderstanding will be narrowed with validity analysis of test items.

Ways I could use help from the group

- Student interview protocol
- Design and statistical correlation analysis of paired or tiered questions; discussion of use of multiple answer questions
- Consensus definition of misconception
- Consensus definition of a concept in biology; description of major genetics concepts
- Additional trials of the GenCI instrument at other universities; possible incorporation of GenCI questions into the BCI.

References Cited

1. Elrod, S.L., 2007, *Genetics Concept Inventory*, Conceptual Assessment in Biology II workshop paper, Boulder, Colorado.
2. Elrod, S.L. and M.M. Somerville, 2007, *Literature-Based Scientific Learning: A Collaboration Model*, Journal of Academic Librarianship, in press.
3. Hott, A.M., C.A. Huether, J.D. McInerney, C. Christianson, R. Fowler, H. Bender, J. Jenkins, A. Wysocki, G. Markle, R. Karp, 2002, *Genetics Content in Introductory Biology Courses for Non-Science Majors: Theory and Practice*, BioScience 52 (11); 1024-1035.
4. Marbach-Ad, G., 2001, *Attempting to break the code in student comprehension of genetic concepts*, J of Bio Educ 35(4):183-189.
5. Marbach-Ad, G. and R. Stavy, 2000, *Students' Cellular and Molecular Explanations of Genetic Phenomena*, J Bio Educ, 34(4): 200-205.
6. Banet, E. and E. Ayuso, 2000, *Teaching Genetics at Secondary School: A Strategy for Teaching about the Location of Inheritance Information*, Sci Ed, 84: 313-351.
7. Wood-Robinson, C., Lewis, J. and Leach, J. 2000 *Young people's understanding of the nature of genetic information in the cells of an organism*, J Bio Educ 35(1): 29-36.
8. Lewis, J., Leach, J. and Wood-Robinson, C. 2000 *Chromosomes: The Missing Link - young people's understanding of mitosis, meiosis and fertilization*, JBio Educ 34(4): 189-199.
9. Lewis, J., Leach, J. and Wood-Robinson, C. 2000 *What's In a Cell? Young peoples understanding of the genetic relationship between cells, within an individual*, J Bio Educ 34(3): 129-132.
10. Lewis, J., Leach, J. and Wood-Robinson, C. 2000 *All In The Genes? Young people's understanding of the nature of genes*, J Bio Educ 34(2): 74-79.
11. Lewis, J. and Wood-Robinson, C. (2000) *Genes, chromosomes, cell division and inheritance - do students see a relationship?*, Int J Sci Educ 22(2): 177-195.
12. Parker, J. *et al*., 2007, *Teaching and Learning Biology at the Undergraduate Level*, Conceptual Assessment in Biology II workshop paper, Boulder, Colorado.
13. Fisher, K.M., 1985, *A Misconception in Biology: Amino Acids and Translation*, J Res Sci Teach 22(1): 53-62.

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.

Connecting K-12 science reform to higher education STEM.
(A workshop with paper.)

Paul J. Kuerbis, Ph.D.
Colorado College
Colorado Springs, Colorado 80903

This paper and workshop discussion will examine how resources originally developed for improving K-12 science have been used successfully for that purpose AND how these tools have the potential for supporting reform in higher education STEM. The author will highlight stories and examples from several projects (K-20) at Colorado College over the past 10 years.

In the 1990s several seminal documents emerged from such groups as the National Academies (of science, medicine and engineering) and the American Association for the Advancement of Science (AAAS) designed primarily to inform and improve K-12 science teaching and learning. Among the reports and books were the AAAS's *Benchmarks* (1993) and *Atlas of Science Literacy* (now in two volumes, 2000, 2007) and the National Academies *National Science Education Standards* (1996), *Inquiry and the National Science Education Standards* (2001), and the often cited *How People Learn (1999)*.

In this paper and the workshop session the focus will be the AAAS *Atlas.* Both volumes help K-20 teachers determine developmentally appropriateness of concepts ("Big Ideas") as well as see how a <u>teaching sequence of concepts</u> might be modified based on the published 'road maps.' These road maps are broadly constructed (e.g., "Flow of Energy Through Ecosystems", "Structure of Matter: atoms and molecules," etc.) and always cut across disciplinary boudaries. Teachers not only see a sensible sequence, they realize how "Big Ideas" from different domains of science are interwoven and interdependent.

In reviewing one map, a chemistry faculty member noted that the conceptual flow map for Structure of Matter is the same sequence he uses in his introductory general chemistry. He discovered this by teaching for 20+ years and now verifies that the road map is a successful pathway with his college students. Yet the portion of the map he analyzed was for grades 6-12. How might maps in the Atlases be used as curriculum planning tools in higher education? This workshop (with paper) will explore answers to this question.

STUDYING C-TOOLS: AUTOMATED GRADING FOR ONLINE CONCEPT MAPS

DOUGLAS B. LUCKIE[1], SCOTT H. HARRISON[2], JOSHUA L. WALLACE[1] AND DIANE EBERT-MAY[3]

[1]Lyman Briggs College of Science and Department of Physiology, [2]Department of Microbiology and Molecular Genetics, [3]Department of Plant Biology, Michigan State University

Abstract. The C-TOOLS project has developing a new assessment tool, the Concept Connector, consisting of a web-based, concept mapping Java applet with automatic scoring. The Concept Connector is designed to enable students in large introductory science classes at the university level to visualize their thinking online and receive immediate formative feedback. The Concept Connector's flexible scoring system, based on tested scoring schemes as well as instructor input, has enabled automatic and immediate online scoring of concept map homework. Criterion concept maps developed by instructors in the C-TOOLS project contain numerous expert-generated or "correct" propositions manually created by expert users connecting two object or subject phrases together with a linking phrase. A range of holistic algorithms as well as WordNet, an electronic lexical database, are being used to test automated methods of scoring. For this study 1298 students created concept maps (with 35404 propositions) were evaluated by automatic grading and for validity of computer generated WordNet® propositions. We studied how successful these approaches were at creating and/or evaluating additional linking words extrapolated from criterion maps generated by experts. By comparing manual assessments of derived propositions to manual assessments of original propositions, the persistence of correctness was evaluated.

Category: Poster Paper

1 Introduction

Expert-level thinking depends on a web of mental connections developed over a lifetime of education and experience (Bruner, 1960). Yet, in an attempt to turn college science students into experts, instructors often just focus on passive transmission of large amounts of "content" in a short time period and then test students to see if they "got it" (NRC, 1999). In response, students tend to focus on practical ways to succeed in their courses and thus often adopt strategies like memorization or rote learning (Ausubel, 1963; Novak & Gowin, 1984). Visual models such as concept maps may help instructors teach expert thinking as well as assess domains of student understanding. In our own learning as scientists, we frequently use visual models (Casti, 1990). The value of knowledge scaffolding tools such as concept maps is that they reveal student understanding about the direct relationships and organization among many concepts.

The use of paper and pencil seems to be the most natural way to create concept maps. Students can easily create shapes, words, lines etc and can add small illustrations. As students become more proficient or engaged in making a concept map, problems arise when they'd like to revise it. Erasing can become tedious and inhibit the process of revision. Using "Post-It" notes can allow easy revision, yet a record or copy of the map is not easily generated in the active classroom. An additional challenge is scoring maps. While grading a single concept map may be less time-consuming than grading a long essay or extended response, it is still more complex than grading multiple choice exams. Even if a chemistry instructor would like to use concept maps in their large introductory course of 500 students, they will point out that grading 500 maps is not practical for them. Computer software is an avenue to address these challenges. In fact, a number of projects, like the Inspiration™ commercial software and the freely downloadable, community-oriented IHMC CmapTools software, present excellent replacements for paper-and-pencil drawing environments and may help engage the resistant student.

Although computer-based tools for concept mapping are available to university faculty, few are web-based and none have embedded assessment components for automated scoring and feedback. The C-TOOLS project is to develop and validate a new assessment tool, the Concept Connector, consisting of a web-based, concept mapping Java applet with automatic scoring and feedback functionality. The Concept Connector is designed to help "make transparent" when students do not understand concepts and motivate students to address these deficiencies. Web-based concept mapping can enable students to save, revisit, reflect upon, share and explore complex problems in a seamless, fluid manner from any internet terminal (Pea et al., 1999). Automated grading and feedback features can allow instructors to use concept mapping on a larger scale.

Automatic grading features associated with the C-TOOLS project's Concept Connector began in 2003 to amplify instructor-designed grading matrices with synonyms from WordNet®[1] (Fellbaum, Ed., 1998). At present,

[1] WordNet is a registered trademark of Princeton University.

for 35404 propositions, 9211 can be evaluated by an automated grading mechanism called Robograder™. WordNet-powered amplification enabled 971 of the 9211 propositions to be evaluated when the existing grading matrices would not otherwise make an assessment. Currently, Robograder™ indiscriminately accepts linking phrase synonyms independent of frequency and word sense.

Visual examinations of automatically graded maps indicated few false positives or false negatives despite Robograder's treatment of multiple synsets as interchangeably equivalent. In theory amplifying grading rubrics by using the superset of all available synonyms should introduce errors into rubrics since multiple and conflicting meanings often exist. One explanation for the observed success of indiscriminate acceptance of synonyms is that users may more likely choose words within a relevant set of synonyms (known in WordNet as a "synset"). Thus, when developing concept maps, users appear to be less inclined to take "stabs in the dark" than compared to a conceptual strategy which favors semantically plausible word choices.

2 Methods

2.1 Goals and Timeline

With both the literature providing a solid theoretical basis for using concept maps and the field of computer science providing the proper software development tools and technology, the C-TOOLS project began in late 2002. A team of faculty from Michigan State University spent much of the first year of the project developing both the Java applet, called the Concept Connector (Figure 1), and the classroom problems sets with concept maps for science students. In parallel with software development was a study of how students use the tool. The Concept Connector was developed through a 'design' experiment (Suter & Frechtling, 2000) that involved testing the tool with undergraduate science-majors in biology, geology, physics and chemistry courses.

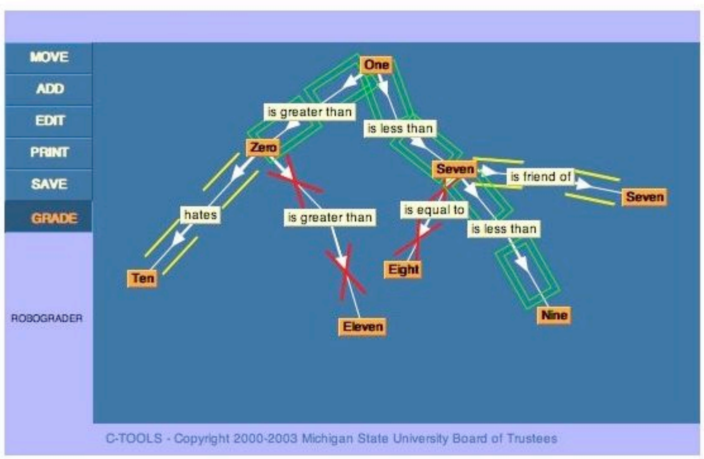


Figure 1: The Concept Connector Java applet graphic user interface (GUI). This particular screenshot shows the Java applet's GUI (blue colored areas), how the software draws a concept map, and how new colors (green and yellow rectangular halos or red X's) appear when the *Robograder* is asked to *GRADE* a concept map (http://ctools.msu.edu).

2.2 Faculty and Students: Concept Mapping in Large Introductory Courses

For the C-TOOLS project, we recruited a cohort of over 1000 freshman and sophomore students enrolled in introductory science-major courses: Biology, Chemistry, and Physic, as well as non-major science courses: Introductory Biology and Geology. During class meetings, students learned how to use the web tools. Students completed concept maps as an integral part of the course. Online concept map-based homework assignments varied widely from analysis of scientific literature to answering a particular homework question. To complete an assignment students typically logged into a website and were presented with instructions and a map space seeded with approximately 10 concepts. The Concept Connector software allows students to move concept words around, organize hierarchy, and add linking words and lines. C-TOOLS exercises often challenged students to first construct a map individually, and submit it to the computer to receive visual feedback. They then could revise the map and resubmit. Finally, they often worked with a partner to complete a final collaborative concept map.

2.3 Holistic Scoring Approaches: Development and Data Analysis

The Concept Connector™ has been created as the combination of an online Java applet that serves as a map drawing tool residing in an HTML page that communicates with server-side software on POSIX-conforming systems such as Mac OS X®, LINUX®, and FreeBSD®. The applet is small in size and is browser-compatible on every OS platform and presents a menu-driven, interactive GUI. In terms of architecture, as a technology, a C-TOOLS server incorporated freely available software tools and followed existing software conventions within the freeware community. By implementing and interacting with necessary software components such as cross-linking databases, resource-specific handlers, and servlets in this manner, the C-TOOLS project was careful to utilize open standards.

The online software allows students to seamlessly create their concept map on an "easel" page, save it in a private "gallery," restore, revise and submit it to receive automatic scoring feedback. In automated grading, our primary goal is to follow the scoring system developed by the Novak group (Novak & Gowin, 1984). "Robograder" gives visual feedback concerning the validity of the semantic relationship between linked words in a proposition. During the study period, automated scoring of student linking words graded 26% of the user-made propositions existing on Michigan State University's C-TOOLS server. In addition to the Novak scoring system, we are studying student maps for interesting trends and testing new "Gestalt" approaches for automated feedback that successfully mimic the human grader (Figure 2).

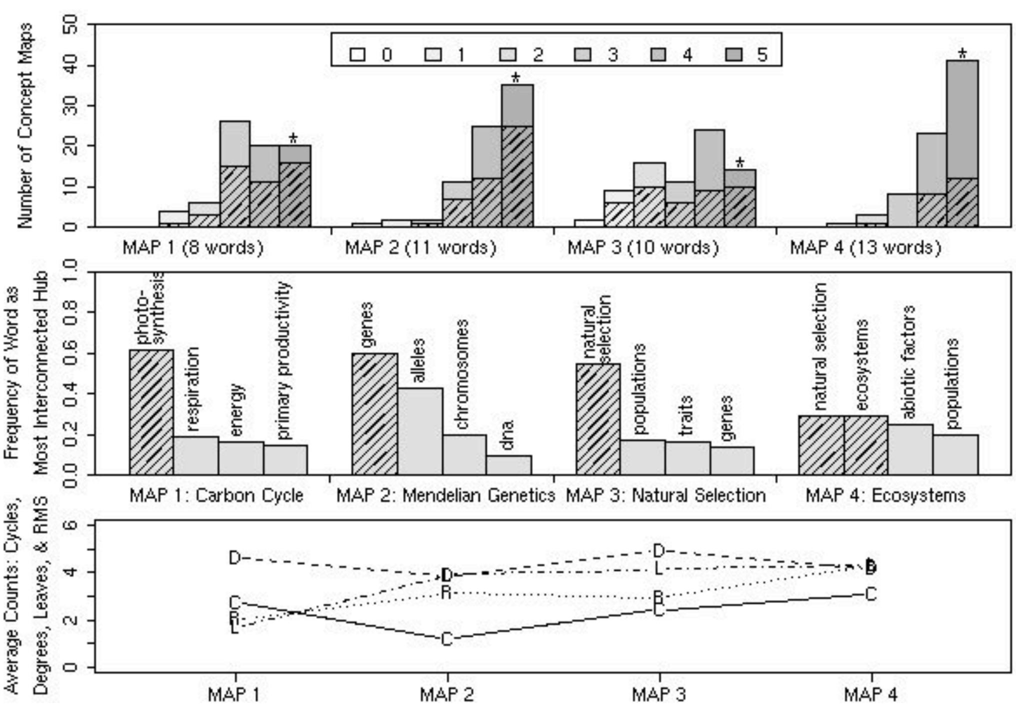


Figure 2. Human expert scoring of student maps from a non-majors biology course (top panel) and software analysis of trends in the map data (lower panels). Panel 1 (top) shows the distribution of scores (graded from 0 to 5) for each of 4 assignments given successively throughout a semester (n=76 students). The striped portions of the bars in panel 1 indicate the distribution of scores for maps that used the top "hub" concept word (for MAP1 this hub word was "photosynthesis," identified in Panel 2 (middle)). Panel 3 shows trends in Gestalt scoring approaches applied to the same maps. These are the average values of 4 network topology measurements for the maps that scored a "5" (*) from each of the 4 assignments. Cycles="C", the number of loops involving 3 or more concept words; Degrees="D", the number of propositions connecting to a given concept word; Leaves="L", the number of terminal ends in the concept map network; RMS, an indicator of non-branching chains within a concept map="R", the root of the mean sum of squared distances between all concept word pairs within a concept map.

We are studying more holistic domains such as frequency of word choice and links, network patterns and evaluative approaches based upon the structure of the map. Figure 2 presents an analysis of data from one C-TOOLS biology course. It aligns the distribution of grades (0-5) given by the expert faculty to student concept maps made during a semester (top panel) with an analysis of most common “hub” concept words found in the student maps (“hub”=concept with most links; middle panel) and “Gestalt” grading strategies where software attempts to evaluate the same student maps via content independent approaches (bottom panel).

C-TOOLS provides a well-curated data source with which to assess trends of classroom learning as shown in Figure 2. The instructor predicted the reduced student performance seen for MAP 3 based on complex interdependencies associated with the “Natural Selection” knowledge domain. The instructor also predicted that those students understanding certain critical concept words, as evidenced in MAP 1 by choosing words such as “photosynthesis” to be the most highly interconnected hub, would score the highest on their concept maps. The shift in grade distribution of maps (top panel, striped portions of bars) using the most popular “hub” word (identified in the middle panel) appears to support the prediction.

Automated “Gestalt” grading approaches currently being tested are based on the network structure of the student concept maps. Methodologies using map network patterns related to hierarchy (“Leaves” and “Degrees”), cross-linking (“Cycles” and “RMS”), as well as the use of software called WordNet® to amplify linking word databases are being studied (Harrison, Wallace, Ebert-May, & Luckie, 2004). In the bottom panel of Figure 2, four automated scoring strategies were tested on student concept maps that received a score of 5. Interestingly, topology measurements termed “RMS” and “Leaves” correlated best with the human grader. The capacity to analyze and verify these predictions will grow in power with the accumulation of additional data and classroom-to-classroom comparisons.

2.4 WordNet Scoring Approaches: Testing and Data Analysis

At the time of this WordNet study there were 35404 propositions available from Michigan State University's C-TOOLS server. A random sample of 250 propositions was gathered and divided into 5 separate sets of 50 each. Each of these original propositions consisted of a starting concept phrase, a linking phrase, and a terminal concept phrase. Manual assessment of propositions was done by hand without aid of electronic references or algorithms. Manual assessments of the 5 sets were performed by the first two authors. Scorings for each proposition were: 1 (correct, e.g. “Photosynthesis - needs - Carbon dioxide”), X (incorrect, e.g. “DNA - transcribes – RNA”), 0 (ambiguous, e.g. “atom - is made of – neutron”), and S (structural violation, e.g. “ocans - evaporation – atmosphere”). Structural violations were for propositions with grammar problems such as spelling errors and linking phrases that do not contain a verb. Ambiguous scores were given to propositions that could only be scored as correct when viewed in a reasonably plausible context of surrounding propositions.

Version 2.0 of the software database WordNet was used to generate "proposition derivatives" by making linking phrase substitutions based on WordNet's thesaurus-like lexical capabilities. There were two criteria for the generation of proposition derivatives. First, derived propositions were made from linking phrases consisting of a single verb. Only 121 of the 250 original propositions met this single verb word criterion. Second, at minimum, the WordNet database had to have three available choices per lexical relationship. An original proposition’s linking verb could thus have up to nine derivatives (i.e. 3 antonyms, 3 troponyms, and 3 synonyms). Each triplet, as generated per lexical relationship (e.g. three antonyms), is called a trio. Trios provide an initial comparative range of data concerning the sense of usage and polysemy counts specific to both lexically similar and dissimilar derivations made from original propositions. With the single verb and triplet criteria, WordNet enabled us to construct 30 antonym derivatives, 243 troponym derivatives, and 234 synonym derivatives per proposition. Grading of proposition derivatives was delegated by the originating proposition sets. Graders A and B both graded derivative set 5. Grader A graded derivative sets 3 and 4. Grader B graded derivative sets 1 and 2.

The manual assessments of original and derivative propositions were scrutinized in order to both summarize and make insights into relationships that may concern automated strategies of assessment. Assessments of original and derived propositions were enumerated in order to show relative ratios of correctness, ambiguity, and grammatical errors. Trios were analyzed for fluctuations in correctness and incorrectness. Trends between correctness and polysemy were investigated.

Graders A and B had reproducible similarity to their scoring patterns as determined by the Kappa statistic (Cohen, 1960). The Kappa statistic (_ = 0.552) was calculated with $p_o = 0.720$ and $p_e = 0.374$ suggesting good reproducibility (0.4 ≤ _ ≤ 0.75). The level of significance for this degree of association is < 0.10. For the manual assessments of the 250 original propositions, 72% of the assessments between the two graders were identical (180 propositions). Opposite assessments of correctness (1 versus X) occurred 5.6% of the time. Remaining differences for the assessment of individual propositions were primarily attributable to issues unrelated to exacting qualifications of correctness. For example, 30 instances of disagreement involved only one grader assigning an S score and 26 instances of disagreement involved one grader cautiously assigning a 0 (ambiguous) score in contrast to 1 or X scorings. While our approach has statistically significant repeatability for scoring ratio properties and strong consistency for exacting qualifications of proposition correctness, further refinement would involve better synchronization between graders' approaches to assumptions of context and handling of grammatical logistics. When looking at the jointly graded WordNet derivative set (n = 144), the agreement between grader A and grader B was 70% (p_o = 0.701). The degree of association is just marginally reproducible based on _ = 0.375 and this reduction may be attributable to fewer shared contextual assumptions between graders due to loss of the original word choice. Scoring dynamics appear to be conserved; joint scorings for derivatives rise in agreement when considering just 1 and X scores, and the _ value does not suggest complete insignificance (_ = 0.13).

Antonym derivatives were found to be always incorrect. Of 21 antonyms graded by grader A, 21 were graded as incorrect. Of the 18 antonyms graded by grader B, 18 were graded as incorrect. Assessments for original, synonym-derived, and troponym-derived propositions are shown in Table 1 and encompass a range of assessment across all four grading categories (1, 0, X, and S). When the range of assessment is limited to 1 and X, grader A found 25.6% of synonym-derived propositions to be correct and 16.8% of troponym-derived propositions to be correct. For 1 and X scorings, grader B found 43.5% of synonym-derived propositions to be correct and 31.7% of troponym-derived propositions to be correct.

Score	Original propositions		Synonym-derived propositions		Troponym-derived propositions	
	Grader A	Grader B	Grader A	Grader B	Grader A	Grader B
Correct	141	123	32	68	21	53
Incorrect	12	32	93	88	104	114
Ambiguous	33	13	16	0	22	0
Structural violation	64	82	0	0	0	1

Table 1: Summary of manual assessment scores for original, synonym-derived and troponym-derived propositions.

The construction of trios involves random sampling from each WordNet-generated set of antonyms, synonyms, and troponyms. If conflicting meanings inside each set cause a general variation of proposition correctness, then clustering of correct or incorrect assessments within trios should not differ from a distribution of correct assessments that is random with respect to triplet structure. For the 57 synonym-derived trios assessed by grader A and the 81 synonym-derived trios assessed by grader B, the distributions showed no significant difference (_^{2} = 1.59, p = 0.66 and _^{2} = 2.07, p = 0.56 respectively). For the 54 troponym-derived trios assessed by grader A and the 71 troponym-derived trios assessed by grader B, the distributions also showed no significant difference (_^{2} = 2.64, p = 0.45 and _^{2} = 5.57, p = 0.13 respectively).

The general variability of correctness occurring within trios was investigated further by measuring how assessment score changes relate to similarities in meaning for derived proposition linking verbs. The WordNet database organizes lexical sets into subsets (termed "synsets") grouped together by similar meaning. Pairs of propositions occurring within trios were analyzed for having dissimilar correctness scores 1 and X, and for whether each proposition's linking verb was a member of the same synset. Shared synset membership for troponym derivatives occurred for 67% (grader A) and 49% (grader B) of all trio pairings that had an assessment score transition from 1 to X. Scoring transitions from 1 to X were next contrasted to within-trio proposition pairs where both propositions were assessed with a score of 1. Shared synset membership for troponym derivative pairs occurred for 100% (grader A) and 81% (grader B) of all such trio pairings that had a common assessment score of 1. Synonym derivatives were analyzed in similar fashion. Shared synset membership for synonym derivatives occurred for 15% (grader A) and 14% (grader B) of all trio pairings that had an assessment transition from 1 to X. Shared synset membership for synonym derivative pairs occurred for 27% (grader A) and 31% (grader B) of all such trio pairings that had a common assessment score of 1. Thus, for both troponyms and synonyms, membership of two verbs in the same synset implicates retained assessments of correctness.

Correctness was then analyzed for its impact on polysemy count distributions. For original propositions, polysemy distribution values were _ = 3.43, _ = 7.72 and _ = 4.08, _ = 6.46 for incorrect and correct propositions respectively. For derived propositions, polysemy distribution values were _ = 6.01, _ = 7.51 and _ = 8.90, _ = 11.40 for incorrect and correct propositions respectively. The increase of polysemy counts with correctness was attributed both to moderately high polysemy count ranges (>20) corresponding to the correctness of propositions by a factor of 2.4 and to a distinct trend for low polysemy count ranges (<5) corresponding to a 10% rise in the incorrectness of propositions.

3 Summary

The C-TOOLS project stems from the combined activities of an interdisciplinary team of faculty from Michigan State University. This National Science Foundation (NSF)-funded project developed a new assessment tool, the Concept Connector, consisting of a web-based, concept mapping Java applet with automatic scoring and feedback functionality. The Concept Connector tool is designed to enable students in large introductory science classes to visualize their thinking online and receive immediate formative feedback. Further details concerning the C-TOOLS project have been previously published (Luckie, Batzli, Harrison & Ebert-May, 2003).

In this study of Holistic and WordNet automated scoring approaches of concept maps, an approach that amplifies correctness across multiple synsets appeared to work on concept maps made by real users. Such an indiscriminating approach was faulty when applied to randomly generated sets of synonyms and troponyms. Thus, the data supports that synonyms and troponyms can be used as sets for further identifying both correct and incorrect propositions. Although it may appear that there is only a 10% gain by using synonyms for automatic grading, this is only from the standpoint of automating the assessment at the proposition level. At the larger concept map level, there are highly interconnected concept words that follow a pattern of classroom consensus and also correspond to student performance (Luckie, Harrison, & Ebert-May, 2004). Better understanding of the linking words around major hubs would aid us to analyze the formative dynamics of how users in a classroom interconnect concepts and, potentially, knowledge domains. Analysis and further improvements to Robograder™ cannot just be limited to synset hierarchies of each individual linking word since there are content-dependent dynamics of semantic overlap that influence how words can sensibly connect to other words (Banerjee & Pedersen, 2003).

References

Ausubel, D. (1963). *The Psychology of Meaningful Verbal Learning*. Grune/Stratton. New York, NY.

Banerjee, S., & Pedersen, T. (2003) *Extended Gloss Overlaps as a Measure of Semantic Relatedness*. Paper presented at IJCAI 2003 – 18th International Joint Conference on Artificial Intelligence.

Bruner, J. (1960). *The Process of Education*. Harvard University Press. Cambridge, MA.

Cañas, A. J., Valerio, A., Lalinde-Pulido, J., Carvalho, M., & Arguedas, M. (2003). *Using WordNet for Word Sense Disambiguation to Support Concept Map Construction*. Paper presented at SPIRE 2003 – 10th International Symposium on String Processing and Information Retrieval.

Casti, J. L. (1990). *Searching for certainty: what scientists can know*. New York, W. Morrow, 496 p.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46.

Collins, A., Joseph, D. & Bielaczyc, K. (2004) Design research: Theoretical and methodological issues. *Journal of the Learning Sciences, 13*(1), 15–42.

Fellbaum, C. Ed. (1998). *WordNet – An Electronic Lexical Database,* MA: MIT Press.

Fisher, K. M. (2000). *SemNet software as an assessment tool*. In J.J. Mintzes, et al (eds.), Assessing science understanding: A human constructivist view. Academic Press. San Diego, CA.

Harrison, S. H., Wallace, J. L., Ebert-May, D., & Luckie, D. B. (2004). C-TOOLS automated grading for online concept maps works well with a little help from "WordNet" Paper presented at CMC 2004 – 1st International Conference on Concept Mapping.

Ihaka, R., & Gentleman R. (1996). R: A Language for Data Analysis and Graphics*, Journal of Computational and Graphical Statistics, 5*, 299-314.

Luckie, D. B., Batzli, J. M., Harrison, S., & Ebert-May, D. (2003). C-TOOLS: Concept-Connector Tools for Online Learning in Science. *International Journal of Learning 10*: 332-338.

Luckie, D., Harrison, S., & Ebert-May, D. (2004). *Introduction to C-TOOLS: Concept Mapping Tools for Online Learning*. Paper in review for CMC 2004 – 1st International Conference on Concept Mapping.

National Research Council. (1999). *Transforming Undergraduate Education in Science, Mathematics, Engineering, and Technology*. National Academy Press. Washington, DC

Novak, J. (1990). Concept Mapping: A Useful Tool for Science Education. *Journal of Research in Science Teaching, 27*(10), 937-949.

Novak, J. D., & Gowin., D. D. (1984). *Learning How to Learn*. Cambridge Press. New York, NY.

Pea, R., Tinker, R., Linn, M., Means, B., Bransford, J., Roschelle, J., His, S., Brophy, S., & Songer, N. (1999). Toward a learning technologies knowledge network. *ETR&D*. *47*(2): 19-38.

Suter, L., & Frechtling, J. (2000). Guiding principles for mathematics and science education research methods. *NSF Report* 00-113.

Wittrock, M. C. (1992). Generative Learning Processes of the Brain. *Educational Psychologist, 27*(4), 531-541.

THE “BIG IDEAS” OF PHYSIOLOGY
PART I: WHAT SHOULD STUDENTS UNDERSTAND?

Joel Michael
Department of Molecular Biophysics and Physiology
Rush Medical College
Chicago, IL
jmichael@rush.edu

Harold Modell
Physiology Educational Research Consortium
Seattle, WA
modell@physiologyeducation.org

Jenny McFarland
Department of Biology
Edmonds Community College
Lynnwood, WA
jmcfarla@edcc.edu

William Cliff
Department of Biology
Niagara University
Niagara, NY
bcliff@niagara.edu

ABSTRACT

The explosion of knowledge in the biological sciences has created a growing problem for biology educators. There is more to know than students can possibly learn. Thus, difficult choices have to be made about we expect students to master. One approach to making the needed decisions is to consider those BIG IDEAS that provide the thinking tools for understanding all biological phenomena. We have identified nine BIG IDEAS which appear to cover all of ***physiology*** and we have begun the process of unpacking them into their constituent component ideas. While such a list does not define the content for a physiology course, it does provide a guideline for selecting the topics on which to focus student attention. This list of BIG IDEAS also offers a starting point for developing an assessment instrument to be used in determining whether students have mastered the important ideas of physiology.

I. INTRODUCTION

The knowledge explosion is alive and well in biology. One of its more visible signs is the length of the textbooks that we recommend to students (see Table 1). The assigned textbook in a typical introductory biology course may be more than 1000 pages long and contain a very large number of ideas, concepts, principles and facts. The situation in physiology is no different, whether we look at human anatomy and physiology books, physiology texts aimed at undergraduates, or medical and graduate level physiology textbooks.

However well the course is taught, students can only learn a fraction of what is in such a book. Furthermore it is clear that they will retain an even smaller fraction over time. Equally important, the focus on learning ever more "content" does not help students understand physiological principles. This is a long-recognized and persistent problem in all of biology (Nelson, 1989; Wright and Klymkowsky, 2005).

TABLE 1
The size of popular, current textbooks of biology and physiology

	Pages of text
INTRODUCTORY BIOLOGY	
Freeman, 2005	1392
Campbell et al., 2004	1312
HUMAN ANATOMY & PHYSIOLOGY	
Saladin, 2007	1248
Marieb, 2007	1296
Martini, 2007	1110
UNDERGRDUATE PHYSIOLOGY	
Sherwood, 2004	801
Widmaier, Raff, Strang, 2004	738
MEDICAL PHYSIOLOGY	
Boron and Boulpaep, 2003	1267
Berne, Levy, Koeppen, Stanton, 2004	978
Guyton and Hall, 2006	1066

What should a student know after having taken a physiology course? Obviously, not everything found in the textbook! What do we want students to retain long after they have completed the physiology course? There is no generally agreed upon answer to these questions.

As we think about these questions, it is useful to reflect on the fact that this is not the case in all disciplines. In physics, for example, there is near universal agreement on what the content of a first level physics course ought to be. In fact, the physics curriculum is remarkably similar across all colleges and universities that offer a degree in physics. It is this nearly universal agreement on what physics students should know that has made it possible for the physics education community to generate assessment instruments with which to determine what students do, in fact, know. The Force Concept Inventory (FCI) developed by Hestenes, Wells, and Swackhamer (1992) was only the first of such concept inventories to be written and widely used.

Can the physiology education community define what students should understand and develop instruments to allow us determine what students do know?

II. CONCEPTUAL ASSESSMENT IN BIOLOGY: AN NSF SPONSORED MEETING

In March, 2007 NSF sponsored a meeting at the University of Colorado in Boulder, CO to consider the issue of conceptual assessment in biology. Twenty-one biology educators from a variety of disciplines came together to discuss the possibility of developing an assessment instrument to determine student understanding of concepts in biology. It is fair to say that the assumed goal was an instrument that would do for biology what the Force Concept Inventory (Hestenes, Wells, and Swackhamer, 1992) has done for physics.

The bulk of this meeting was devoted to attempts to decide what it is that students should be expected to know. We need to know what should be assessed before deciding how to generate an assessment instrument.

However, it is difficult to decide what is meant by a “concept.” An informal survey (Michael, unpublished results) of at least a dozen physics educators presenting papers at an American Association of Physics Teachers meeting resulted in no operationally useful definition of what they meant by the term “concept” when referring to the Force Concept Inventory. Furthermore, it seems likely that the term means something different to biologists than it does to physicists.

Participants did find the notion of “BIG IDEAS” to be a useful construct with which to attack the problem of defining what to assess. Duschl, Schweingruber and Shouse (2007) offer the following definition of this term:

> Each [BIG IDEA] is well tested, validated, and absolutely central to the discipline. Each integrates many different findings and has exceptionally broad explanatory scope. Each is the source of coherence for many key concepts, principles and even other theories in the discipline.

There was considerable discussion of what BIG IDEAS might best represent the bases for the biological sciences, and a tentative list was eventually generated. But it is not clear that this list will serve all of the various biological disciplines equally well, and continued consultation among biology educators will be needed to arrive at a definitive set of BIG IDEAS.

A brief report of this meeting will appear in *Advances in Physiology Education* (Michael, in press).

Other conversations about disciplinary big ideas have been occurring at many levels in the past several years. The Washington State Board of Community and Technical Colleges has convened two workshops (in 2003 and 2007) to discuss disciplinary big ideas. The central question they addressed was "Assuming for the moment that [an] introductory course is the LAST course the students will ever take in the area, what core discipline-specific ideas/concepts do you want ALL students taking the course to understand deeply; i.e., be able to apply and even transfer to other settings five years later?" Biologists participating in these meetings generated lists of big ideas that closely mirror the list developed at the CAB meeting.

III. BIG IDEAS IN PHYSIOLOGY: WHAT STUDENTS SHOULD KNOW

As physiology educators we were interested in generating a set of BIG IDEAS in ***physiology*** that could inform our teaching and that of our colleagues. The nine BIG IDEAS described here appear to be able to serve as the foundation for physiology. This list is not definitive and input from the entire physiology education community is needed if we are to reach a broad consensus about what ideas students should learn.

Each of the BIG IDEAS is defined and its context within physiology is described. An example of important physiological phenomena, commonly taught in physiology courses at all levels, is provided for each BIG IDEA.

Feder (2005) has proposed a set of "central core ideas or concepts that an undergraduate education might strive to communicate," an agenda not unlike our agenda. There is considerable overlap between his list and our list of BIG IDEAS.

We believe that course objectives in our physiology courses should reflect and support the learning of these disciplinary BIG IDEAS but these are not course objectives. ***This list is NOT a prescription for the content of a physiology course***, but reflects concepts that are at the core of our discipline.

BIG IDEA I

Living organisms are machines whose ***causal mechanisms*** can be understood by applications of the laws of physics and chemistry.

In some sense this BIG IDEA is a refutation of the notion of vitalism that has never completely disappeared from our culture. If this is all that it describes it would be better to think of it as a description of the nature of the research enterprise in the biological sciences.

It is, however, something more than this. It is essential that students recognize that understanding physiological systems (being able to explain the mechanisms producing a response or predicting the occurrence of responses) requires the ability to think causally. Physiology teachers believe that this requirement is one of the major sources of the difficulties that students having in learning physiology (Michael, 2007). In particular, students have difficulty distinguishing between cause and effect (does lung volume change cause pressure change or visa versa).

There are other implications that must also be considered. The properties (states) and functions of the organism are measurable, and changes in the measured values are meaningful. Physiology is thus, at least partially, a quantitative discipline, and the learner must pay attention to units of measurement and to orders of magnitudes of measured variables.

Finally, this BIG IDEA is an antidote to the kinds of teleological thinking that are so prevalent among students and others.

EXAMPLE: Blood flow to exercising muscle is increased. This is a consequence of the muscle's increased metabolism generating local stimuli that relax arteriolar vascular smooth muscle and reducing resistance to flow. (Students commonly argue that blood flow increases because the exercising muscle "needs" more oxygen, without recognizing that "need" is not a mechanism.)

BIG IDEA II

The cell is the smallest, self-replicating unit of integrated function . The organism is made up of tissues comprised of different cells with specialized structures and functions.

This BIG IDEA is one of the oldest in the "modern" era of biology. It is so elemental that it is usually implicitly assumed, not explicitly stated. As a result the important consequences that follow from it are often unappreciated.

The cell membrane that separates the interior of the cell from the external environment has specific properties, and these contribute to the specialized functions of every cell. In a complex, multicellular organism cells have specialized functions, with no one cell able to perform all of the tasks required to maintain the organism.

EXAMPLE: The islet of Langerhans in the pancreas is comprised of three different types of cells, each of which releases a different hormone involved in the regulation of glucose metabolism.

BIG IDEA III

Life requires ***information flow*** in and between cells and between the environment and the organism.

Information is one of those terms that is frequently used in everyday discourse, although its meaning in that context may not always correspond to its technical meaning. Information flow is present at multiple levels in every organism and is, in fact, one of the hallmarks of living systems.

Genetic information determines, in complex ways, the structure and function of the organism as it develops from a fertilized egg. Information about the state of the external world must be available to allow appropriate responses to the many conditions that pose a danger to the organism. Information must be passed from cell to cell in order to make possible the coordinated responses of the organism to changes in both the internal and the external environment.

EXAMPLE: The strength of contraction of a skeletal muscle, which must be matched to the task to be performed, is determined by information delivered to the muscle by the number of active neurons and the frequency of firing action potentials in the nerve innervating the muscle.

BIG IDEA IV

Living organisms must obtain matter and energy from the external world to continue to exist. That ***matter and energy must be transferred and transformed*** in a varied of ways in order to build the organism and to perform work (from the cellular to the organismal levels).

All functions of living organisms are energy dependent and all organisms must have access to energy in order to survive (plants from sunlight and animals from plants or other animals). Energy in the form of compounds with high-energy bonds is used to synthesize biological molecule, to power solute pumps, and to produce contraction of muscles.

Regulation and control (components of the BIG IDEA of homeostasis) involves altering the function of cells by altering their uses of matter and energy.

EXAMPLE: The distribution of solutes across the cell membrane is created and maintained by pumps in the cell membrane that move solutes against their electrochemical gradient. The work to accomplish this comes from the release of energy stored in ATP molecules.

BIG IDEA V

Homeostasis is a process that maintains the internal environment of living systems in a more or less constant state.

This is perhaps the defining BIG IDEA in physiology.

Important system parameters are measured and the measured values are compared to a pre-determined set-point, or desired, values (although we do not know the mechanisms generating all of these set-points). The difference is used to generate signals (information) that alter the functions of the organism to return the regulated variable towards its pre-set determined value.

EXAMPLE: In mammals body temperature is maintained more or less constant in the face of changes to environmental temperature by manipulating heat production and heat loss through various mechanisms.

BIG IDEA VI

To understand the behavior of the organism requires understanding the relationship between the ***structure and function*** of the organism, since function is dependent on structure and structure must match the functional needs of the organism.

This BIG IDEA is, on one level, a fairly abstract statement of the obvious interaction between the way in which the pieces of a mechanism are assembled into a system and the functions that the system can carry out. However, it also describes several very specific examples of commonalities that extend across many different physiological systems. For example, when two systems carry out similar functions certain features of their structure can be expected to be similar.

EXAMPLE: Gas exchange in the lungs and absorption of the products of digestion in the small intestine occur, in the latter case in part, by the process of diffusion. In both cases, the area across which diffusion occurs is very large and the distance to be traversed is short as a consequence of the structure of the respective systems.

BIG IDEA VII

Living organisms carry out functions at many different ***levels of organization*** simultaneously.

Research in physiology currently extends across levels of organization that include: molecules, cell components, whole cells, tissues, organs, organ systems, and the whole organism.

At each level we encounter emergent properties that can not simply be accounted for by any simple "summation" of properties at lower levels.

EXAMPLE: Knowing the properties of each isolated component involved in blood pressure regulation does not allow one to predict the behavior of the negative feedback system that is the baroreceptor reflex.

BIG IDEA IIX

All life exists within an ***<u>ecosystem</u>*** comprised of the physicochemical world and the total biological world.

Physiology is not typically taught from an ecological or even environmental standpoint, with the possible exception of comparative physiology. Nevertheless, it is clear that the individual organism exists, and survives to reproduce or not, as part of an ecological system. Comparative physiology clearly applies this BIG IDEA in significant ways, and more attention to this is undoubtedly warranted in the general physiology education community.

EXAMPLE: A number of industrial chemicals (DDT, PCB) and plant products have estrogen-like properties that can disrupt the body's reproductive functions.

BIG IDEA IX

<u>Evolution</u> provides a scientific explanation for the history of life on Earth and the mechanisms (at the molecular level and at the level of species etc) by which changes have occurred to life.

Over the past 100 or so years, this BIG IDEA has become the major organizing idea for essentially all aspects of biology. Its implications inform all biological sciences, although the teaching of these sciences draws upon the explanatory power of the BIG IDEA of evolution to varying degrees. Explanations of physiological phenomena do not commonly invoke the processes of evolution, although this is more common in studies of comparative physiology.

EXAMPLE: The hemoglobin mutation that results in HbS confers some protection against malaria (a useful evolutionary adaptation), although it also results in serious illness resulting for impaired tissue perfusion.

Table 2 contains a list of these BIG IDEAS in summary form.

TABLE 2
Big Ideas In physiology

I. Living organisms are ***causal mechanisms*** whose functions are to be understood by applications of the laws of physics and chemistry.
II. ***The cell*** is the basic unit of life.
III. Life requires ***information flow*** within and between cells and between the environment and the organism.
IV. Living organisms must obtain matter and energy from the external world. This ***matter and energy must be transformed and transferred*** in varied ways to build the organism and to perform work.
V. ***Homeostasis*** (and "stability" in a more general sense) maintains the internal environment in a more or less constant state compatible with life.
VI. Understanding the behavior of the organism requires understanding the relationship between ***structure and function*** (at each and every level of organization).
VII. Living organisms carry out functions at many different ***levels of organization*** simultaneously.
IIX. ***Evolution*** provides a scientific explanation for the history of life on Earth and the mechanisms by which changes to life have occurred.
IX. All life exists within an ***ecosystem*** made up of the physicochemical and biological worlds.

It is important to emphasize that this list of BIG IDEAS in phyiology is not to be read as defining the content of a course or a curriculum. It is a description of the ideas that biologists use in attempting to make sense of biological phenomena. It is list a list of ideas that that should be present in a physiology course in varying proportions depending on the specific subject matter of the course. The relationship between the list of BIG IDEAS and the content of courses or curricula will also vary amongst the different biology disciplines.

The explanatory power of each of these BIG IDEAS for understanding physiology varies considerably. There can be no question that ***homeostasis*** is THE central idea in physiology, while for most (non-comparative) physiologist ***ecosystems*** play little role in helping to organize their thinking. Finally, we need to distinguish between the uses of these BIG IDEAS in doing physiology research and their use in teaching physiology.

IV. UNPACKING THE BIG IDEAS OF PHYSIOLOGY

Like atoms which can be unpacked into a great many smaller particles, each BIG IDEA is made up of a collection of other ideas that may be "smaller" in scope, but nevertheless, have deep and wide explanatory power. We are calling these "component ideas."

What follows is an attempt to unpack the BIG IDEAS most important in <u>physiology into their component ideas</u>. There is still much to be done to complete this process and input from interested physiologists will be most helpful.

BIG IDEA: I. CAUSAL MECHANISM

component: (1) The laws of physics and chemistry describe the functioning of the organism

component: (2) The organism is a "mechanism" in which changes in function arise from the behavior of the mechanism and in which changes "propagate" to affect other functions

component: (3) States and functions of the organism are quantifiable and the absolute magnitudes and changes in magnitude are important to understanding the system

BIG IDEA: II. THE CELL

component: (1) The cell membrane contains the contents of the cell and determines what can enter and leave the cell

component: (2) The internal constituents and state of the cell are different than the extracellular environment

component (3) Although all cells have the same DNA not all genes are expressed in every cell

component: (4) As a consequence, cells have many common functions, but also many specialized functions

component: (5) The organism is a collection of cooperating cells, each cell type contributing its special functions to the "economy" of the organism

BIG IDEA: III. HOMEOSTASIS

component: (1) The organism attempts to maintain a more or less constant internal environment that is different than the external environment

component: (2) Stability of the internal environment occurs via information flow in the form of negative feedback

component: (3) Some limited set of internal system parameters are regulated (held more or less constant) by the manipulation of other parameters whose values are controlled component: (4) The “desired” value of a regulated parameter behaves like a “set-point”

component: (5) The value of the set-point can change as the situation of the organism changes

component: (6) The actual value of a regulated variable must be measured by the body (a parameter can only be regulated if it can be measured)

component: (7) The determinants of a regulated variable must be controlled by the body by altering matter/energy transformations

BIG IDEA: IV. INFORMATION FLOW

component: (1) Transmission of genetic information
- (a) Genetic information is coded in DNA making up genes
- (b) Expression of a gene (reading of the code) results in the cell producing a protein (enzyme)
- (c) Expression of genetic information can be turned on and off leading to cell differentiation
- (d) Expression of genetic information determines intracellular function

component: (2) Neural information processing
- (a) Information is encoded and transmitted by all-or-non action potentials generated in neurons and sensory receptors
- (b) Information is passed from neuron to neuron by chemical transmission at synapses, some of which are excitatory and some of which are inhibitory
- (c) The probability of a neuron firing is determined by the balance between the excitatory and inhibitory inputs

component: (3) Chemical information processing
- (a) Cells produce and release signaling molecules which affect their own function and the function of neighboring cells
- (b) Endocrine cells produce and release hormones which are carried to all cells in the body by the circulation
- (c) In order to respond to a signaling molecule a cell must have a specific receptor for that molecule
- (d) When signal molecules bind to a receptor they alter target cell function by altering intracellular enzyme activity

BIG IDEA: V. MATTER/ENERGY TRANSFER AND TRANSFORMATIONS

component: (1) Many physiological processes affect and are affected by changes in the equilibrium state of intra- and extracellular chemical reactions
component: (2) Solutes move across a membrane either passively (down an electrochemical gradient) or actively (using metabolic energy to power a pump)
component: (3) Flow (bulk flow, diffusion) of a substance occurs as the result of an energy gradient
component: (4) Energy is stored in high energy bonds in the constituent molecules of biological systems
component: (5) This energy is used in biosynthesis, moving solutes, and powering muscles

BIG IDEA: VI. STRUCTURE/FUNCTION RELATIONSHIPS

component: (1) The 3-D structure of cells and tissues is a determinant of the functions of the cell and tissue
component: (2) Surface area is a determinant of the movement of all substances and hence surface area (and the surface to volume ratio) is a determinant of function
component: (3) All physical objects (cells, tissues, organs) have elastic properties that are determinants of function

BIG IDEA: VII. LEVELS OF ORGANIZATION

component: (1) Biological organisms function at many levels of organization (atoms to the whole organism) that exist on different physical scales.
(2) Processes occurring on one levels can often be explained by mechanisms occurring at lower levels (reductionism)
(3) Some phenomena at a particular level of organization can not be fully explained by mechanisms occurring at lower levels; such emergent properties represent more than the "sum" of mechanisms at lower levels

Unpacking the BIG IDEAS in/for other biological disciplines (perhaps even for different courses in the same discipline) will likely yield a different list of component ideas than the one presented here.

V. WHAT DO WE DO WITH A LIST OF BIG IDEAS?

The first thing that can be done with a list of BIG IDEAS and their component ideas is to make decisions about what we want students to be able to do and understand, and about which BIG IDEAS contribute to students reaching those goals. To say that all students should understand the BIG IDEAS is not to say that is ALL they need to understand. But in the finite time we have for student learning in any course, we must start to make decisions about what is more important than something else. Although these BID IDEAS should not be taken as a list of topics to be covered in any physiology course, they should be taken into consideration in determining course objectives at every educational level.

For example, what do we want students to understand about the respiratory system? The importance of the BIG IDEA of homeostasis suggests that students need some understanding of the regulation of arterial PO_2 and PCO_2 by the system. That means they need to understand that there are neural receptors measuring both variables and that both variables can be changed by altering alveolar ventilation. Do students need to understand the differences in the properties of the central PCO_2 receptors and the peripheral receptors that measure both PO_2 and PCO_2? That will depend on the overall goals of the particular course we are talking about. Do students need to understand the consequences of a ventilation/perfusion imbalance in determining the values for arterial PO_2 and PCO_2? Again, that depends on the students and course.

With more known about the physiological mechanisms of the body than can possibly be learned, a principled approach to deciding what to include in the course is of some considerable benefit. Students also need to know that what we expect from them is understanding of the BIG IDEAS and the application of them, and that this is more valuable than knowing a hundred isolated facts.

The other thing that we can do with a list of BIG IDEAS is to generate an assessment tool with we can determine whether students do, in fact, understand what we expect them to understand. Assessing student understanding of BIG IDEAS can be done independently of their knowledge of and understanding of the details of particular physiological systems. Such an assessment instrument would then allow us to: (1) measure individual student learning, (2) determine the success of our course in helping students learn, and (3) determine the efficacy of new, experimental interventions to promote learning with understanding. Another positive consequence of the use of such an assessment instrument is that students will begin to believe that these BIG IDEAS are important. Students pay attention to, and take seriously, that which is assessed.

In the following paper we will discuss what such an assessment might look like and how we can proceed to develop it.

REFERENCES

1. Berne, R. M., Levy, M. N., Koeppen, B. M., and Stanton, B. A. (editors). (2004). *Physiology*, 5th edition. St. Louis, MO: Mosby.

2. Boron, W. F. and Boulpaep, E. L. (Editors). (2003). *Medical physiology*. Philadelphia, PA: Saunders.

3. Campbell, N. A., Reece, J. B., Molles, M., Urry, L. A., and Heyden, R. (2005). *Biology*, 7th edition. San Francisco: Benjamin Cummings.

4. Duschl, R. A., Schweingruber, H. A. and Shouse, A. W. (Editors). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: the National Academies Press.

5. Feder, M. E. (2005). Aims of undergraduate physiology education: a view from the University of Chicago. *Advances in Physiology Education*, **29**, 3-10.

6. Freeman, S. (2005). *Biological science*, 2nd edition. San Francisco: Benjamin Cummings.

7. Guyton, A. C. and Hall, J. E. (2006). *Textbook of medical physiology*, 11th edition. Philadelphia, PA: Elsevier Saunders.

8. Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, **30**, 141-158.

9. Marieb, E. N. and Hoehn, K. (2007). *Human anatomy and physiology*, 7th edition. San Francisco: Benjamin Cummings.

10. Martini, F. H. (2007). *Fundamentals of anatomy & physiology*, 7th edition. San Francisco: Benjamin Cummings.

11. Michael, J. (2007). What makes physiology hard for students to learn? Results of a faculty survey. *Advances in Physiology Education*, 31, 34-40.

12. Michael, J. (in press). Conceptual assessment in the biological sciences: A National Science Foundation-sponsored workshop. *Advances in Physiology Education.*

13. Nelson, C. (1989). Skewered on the unicorn's horn: The illusion of tragic tradeoff between content and critical thinking in the teaching of science. In L. Crow (editor). *Enhancing critical thinking in the sciences*. Washington, DC: Society for College Science Teaching. Pp. 17-27.

13. Saladin, K. (2007). *Anatomy & Physiology: The unity of form and function*, 4th edition. McGraw-Hill.

14. Wright, R. L. and Klymkowsky, M. W. (Fall, 2005). Points of view: Content versus process: Is this a fair choice? *Cell Biology Education*, 4, 189-198.

THE "BIG IDEAS" OF PHYSIOLOGY
PART 2: HOW DO WE KNOW IF THEY UNDERSTAND THEM?

Joel Michael
Department of Molecular Biophysics and Physiology
Rush Medical College
Chicago, IL
jmichael@rush.edu

Harold Modell
Physiology Educational Research Consortium
Seattle, WA
modell@physiologyeducation.org

Jenny McFarland
Department of Biology
Edmonds Community College
Lynnwood, WA
jmcfarla@edcc.edu

William Cliff
Department of Biology
Niagara University
Niagara, NY
bcliff@niagara.edu

ABSTRACT

If we expect students to understand the BIG IDEAS of ***physiology*** it is essential that we be able to assess whether they have achieved the level of understanding expected. We describe a process for generating an assessment instrument that can be used with students at all post-secondary levels. This process starts with asking open-ended questions to elicit samples of student thinking about the physiological issues raised. Distracters for multiple choice items are then generated from these free responses. This assessment instrument (the Physiology Big Idea Inventory) can be used to ask whether individual students have achieved conceptual mastery of physiology, whether a physiology course is successful in helping students to learn physiology, or whether some experimental educational treatment works better than some other treatment to improve student learning.

I. INTRODUCTION

In the previous paper (Michael, Modell, McFarland, and Cliff, 2007) we defined nine BIG IDEAS and their component ideas that underpin physiology. One goal for any physiology course should be student understanding of these BIG IDEAS.

How do we assess whether students do, in fact, understand the BIG IDEAS? What is needed is an assessment instrument that specifically targets the student's ability to use an understanding of a BIG IDEA to think about a physiological phenomenon, at least in part, independently of their knowledge of the facts about that phenomenon.

Such a physiology assessment instrument could serve the same functions for the physiology education community that the Force Concept Inventory (Hestenes, Wells, and Swackhamer, 1992) serves for the physics education community. The FCI is thought to test students' understanding of the concepts of Newton's law of motion independently of the students' ability to solve quantitative problems about motion. The FCI has been used to assess individual student understanding of the concepts of physics, the success of courses in helping students learn these concepts, and as a tool for physics education research (see Hake, 1998).

Following development of the FCI and its successful use, a number of other similar inventories have been written in physics (see Maloney, O'Kuma, Hieggelke, Van Heuvelen, 2001), in engineering (see Evans et al., 2003), in chemistry (see Pavelich et al., 2004), and in geoscience (Libarkin and Anderson, 2006).

In this paper we describe a process for developing a Physiology Big Idea Inventory and suggest several approaches to writing questions with the goal of producing a single assessment instrument appropriate for testing all post-secondary students of physiology. Such an assessment tool would contribute to efforts to reform physiology education at all levels.

II. BIG IDEAS OF PHYSIOLOGY

Table 1, taken from our previous paper (Michael, Modell, McFarland, and Cliff, 2007), summarizes the nine BIG IDEAS that students should understand. A more detailed description of them, and the component ideas into which they can be unpacked, can be found in that paper.

It is important to recognize that testing the students' understanding of these BIG IDEAS is difficult to do directly. That is, it is difficult to pose questions about the BIG IDEAS themselves that do not simply elicit the recall of memorized definitions acquired from the lectures or textbooks. However, as we will show, it is possible to write questions dealing with the component ideas that are unpacked from the BIG IDEAS.

TABLE 1
Big Ideas In physiology
(from Michael et al., 2007)

I. Living organisms are ***causal mechanisms*** whose functions are to be understood by applications of the laws of physics and chemistry.
II. ***The cell*** is the basic unit of life.
III. Life requires ***information flow*** within and between cells and between the environment and the organism.
IV. Living organisms must obtain matter and energy from the external world. This ***matter and energy must be transformed and transferred*** in varied ways to build the organism and to perform work.
V. ***Homeostasis*** (and "stability" in a more general sense) maintains the internal environment in a more or less constant state compatible with life.
VI. Understanding the behavior of the organism requires understanding the relationship between ***structure and function*** (at each and every level of organization).
VII. Living organisms carry out functions at many different ***levels of organization*** simultaneously.
IIX. All life exists within an ***ecosystem*** made up of the physicochemical and biological worlds.
IX. ***Evolution*** provides a scientific explanation for the history of life on Earth and the mechanisms by which changes to life have occurred.

III. DEVELOPING AN ASSESSMENT INSTRUMENT FOR PHYSIOLOGY

In thinking about creating a physiology inventory, it is useful to begin by considering some of the features of the Force Concept Inventory (Halloun and Hestenes, 1985; Hestenes, Wells, and Swackhamer, 1992). The FCI is a set of multiple choice questions each of which describes a scenario, poses a question about that scenario, and requires students to apply their understanding of the underlying concepts to select an answer. Such questions are clearly different in style and content from typical physics course exam questions which require some analysis of the scenario described, the identification of relevant equations, manipulation of the equations, and the calculation of a numerical value. The questions making up the FCI are unquestionably testing something different than what is tested by the usual course exam questions.

One significant feature of the FCI is that the distracters, the wrong answers, were written to reflect common student misconceptions (Halloun and Hestenes, 1985). Thus, performance on the FCI assesses both student understanding of the concepts (correct answers) and the presence of misconceptions (wrong answers).

Another important feature of the FCI is that the questions that make it up are appropriate for physics students at any post-secondary educational level. This is, in part, the consequence of physics being a discipline with an essentially universally agreed upon curriculum. It is also a consequence of the focus of the FCI being on qualitative prediction based on an understanding of Newton's Laws, not the analytic or numerical solution to problems.

Physiology is quite different from physics in several regards. For example, in physiology it is not obvious what would distinguish typical exam questions from conceptual questions testing a student's understanding of the BIG IDEAS of physiology. The distinction between testing quantitative/analytical abilities and conceptual understanding is not a prominent feature in physiology.

Equally important, although there is a fairly standard curriculum for physiology courses at the community college level (driven in part by pre-nursing course requirements), other introductory level physiology courses can differ significantly in their content. It is thus an open question whether we can write a set of assessment items that can be used at all post-secondary academic levels.

Thus if the goal is a universally applicable concept inventory the questions must be relatively content independent, or at least deal with content likely to be found in all courses, and the items must be written in simple language.

It seems reasonable to assert that physiology conceptual questions ought to require only an understanding of the BIG IDEAS (and their unpacked component ideas) independently of their knowledge of the details of a physiological system. If we want to know if a student understands the BIG IDEA of homeostasis we do not want the students' ability to correctly answer the question to be determined by his/her remembering the difference between the carotid baroreceptor and the aortic baroreceptor, or their remembering that the central chemorecpetors make a greater contribution to respiratory drive than do the peripheral chemoreceptors. Of course, this does not mean that the instructor can not also expect that students will understand these differences.

Similarly, we do not want the students' ability to answer these questions to be dependent on their having mastered the often esoteric terminology, jargon, and acronyms that abound in physiology textbooks, although it is appropriate to expect some facility with the fundamental "language" of physiology.

These requirements pose significant challenges to writing assessment items. To make questions "course content independent" we can write questions based on topics that are likely to be present in a majority of course. However, the resulting question bank will most likely contain questions that are not appropriate for some courses. Another approach is to write questions which contain all the "facts" that are needed to answer the question ***given an understanding of the BIG IDEAS involved***.

Such questions could be base on:

(1) common, everyday situations such as exercise, exposure to high altitude, development of a fever (see Michael, 1998);
(2) "imaginary" animals with specifically defined characteristics (see Question 3 below); and
(3) non-mammalian species with which most students will be unfamiliar.

With each type of question the student must apply his or her understanding of the BIG IDEAS of physiology to selecting an answer.

If it is not possible to develop one inventory that is universally applicable, then at the least, we want to develop a set of resources that will make it possible to develop assessment instruments tailored to particular educational levels.

The procedures to be followed in developing a reliable and valid conceptual inventory are well known and have been described for a number of different disciplines: physics, geoscience (Libarkin and Anderson, 2007), biology (Garvin-Doxas and Klymkowsky, 2007). Physiology misconception inventories have been developed in a similar way (Michael et al., 1999; Michael et al., 2002). Basically the steps to be followed consist of:

- Writing open-end questions about BIG IDEAS to be administered to a large, heterogenous sample of students.
- Use student answers to these questions to generate distracters for multiple choice questions (MCQ's).
- Have the MCQ's reviewed by content experts and experienced physiology teachers.
- Pilot the vetted MCQ's with large, diverse student populations.
- Revise the questions as needed and pilot test them again.
- Conduct focused interviews with a small sample of students to validate that the distracters used reflect student thinking.
- Revise the MCQ's as needed.
- Use the assessment items with large numbers of students and collect and achive results.

IV. EXAMPLES OF OPEN-ENDED QUESTIONS ABOUT THE BIG IDEAS OF PHYSIOLOGY

What follows are some very preliminary attempts to generate open-ended questions to be used as described above. Such questions must encourage students to respond in a very open fashion, but must also be sufficiently focused so that students stay, more or less, on task. These are intended to be answered by no more than a short paragraph.

Question 1

Muscle cells and nerve cells perform different functions in the body, yet both require the availability of oxygen in order to maintain their normal functions. Explain.

This question tests the students' understanding of the need for energy to power all the functions of all cells.

BIG IDEA: V(5)

Question 2

Muscle and nerve cells clearly have different functions in the body although they have exactly the same DNA in their nucleus. How do different functions arise in cells with the same DNA?

Although the genes are the same, different sets of genes are expressed (ie., some are NOT expressed) leading to the development of different structures and different functions.

BIG IDEA: IV(1)

Question 3

Investigators observe that in spite of changes in activity and in the external environment, Tribbles exhibit a nearly constant concentration of X in their blood. What does this tell you about Tribbles' physiological mechanisms and substance X?

It can be concluded that the concentration of X is measured by some system in the body, and that alterations in body function can bring about changes in the concentration of X .

BIG IDEAS: III(6 & 7)

Question 4

How do the structural characteristics of the lungs supports their function in gas exchange between the atmosphere and the blood? What structural properties would you predict are present in the lungs to enable this function?

To maximize exchange by diffusion a big surface area and short distance for diffusion are needed.

BIG IDEA: VI(2)

Question 5

The loss of fluid (filtration) from glomerular capillary (in the kidneys) is much greater than in skeletal muscle capillaries. How do the two types of capillaries differ from each other? Explain.

There is a greater area of pores in glomerular capillaries than in skeletal muscle capillaries.

BIG IDEA: VI(2)

Question 6

Very small changes in the concentration of a hormone H in the blood can result in a very large change in the output X from some tissue. Explain the mechanism that accounts for this.

Hormones carry information (only a very small concentration is required) which alters cell function.

BIG IDEA: IV(3)

Question 7

Respiration increases (amount of air moved per minute) when an individual exercises. What causes this to happen?

Mechanisms are activated to make available increased oxygen to exercising muscle. The response does not occur because the muscles “need” more oxygen.

BIG IDEA: I(2)

Question 8

Muscle contain enough high energy compounds to produce ATP to power muscle contraction for only a very short time. Explain how a marathoner is able to run for 2+ hours.

Other cells make glucose which is then supplied to the exercising muscle cell.

BIG IDEA: II(5)

Question 9

Sucrose is a sugar that can not cross the walls of the capillaries. If sucrose is introduced into the circulation water leaves the tissues and enters the circulation. Explain.

An osmotic pressure gradient is presents and results in movement of water down its own concentration gradient.

BIG IDEA: V(2)

Question 10

The XYZ cell of a newly discovered species is found to have a sodium concentration inside the cell of 200 mM/L and a sodium concentration outside the cell of 20 mM/L. What can you conclude about the membrane of the XYZ cell?

The cell membrane is either impermeable to sodium or there is a “pump” in the membrane that actively maintains the concentration gradient for sodium.

BIG IDEA: II(1 & 2)

These questions can be answered by students at essentially any academic level. From written answers we will extract, using student language, distracters to be used in multiple choice questions. These will then be piloted by administration to a large, heterogeneous student population.

V. WHAT DO WE DO NEXT?

We are developing a bank of open-ended questions that test all of the BIG IDEAS through an assessment of student understanding of their component ideas. These questions will be circulated to physiology educators (members of the American Physiological Society and the Human Anatomy and Physiology Society) for comments and corrections. Answers to these questions from the largest and most diverse group of students will be collected using a web-based system. With these responses we will begin the process of constructing a multiple choice assessment instrument following the steps outlined above.

VI. CONCLUSION

Our goal is the development of an assessment instrument appropriate for use in all physiology courses to be used to determine whether students understand (can apply) the BIG IDEAS of physiology. This will necessitate developing questions that are written in non-technical language, and require minimal detailed content knowledge to be answered. Questions testing each of the BIG IDEAS and their component ideas (Michael, Modell, McFarland, and Cliff, 2007) will be written, allowing instructors to choose which questions are most appropriate for use in their course. Even if it proves impossible to produce questions that are useable at all educational levels, we believe that we will have produced a resource bank with we can tailor assessment instruments for different courses.

REFERENCES

1. Duschl, R. A., Schweingruber, H. A. and Shouse, A. W. (Editors). (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: the National Academies Press.

2. Evans, D. L. (moderator). (2003). Progress on concept inventory assessment tools. Proceedings of the 33rd ASEE/IEEE Frontiers in Education Conference, Boulder, CO. pp. T4G-1 – T4G1-8.

4. Garvin-Doxas, K. and Klymkowsky, M. (2007). Building the biology concept inventory. Unpublished manuscript that can be found at http://bioliteracy.net/CABS%202007.html.

5. Hake, R. R. (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.

6. Halloun, I. A. and Hestenes, D. (1985). Common sense concepts about motion. *American Journal of Physics*, **53**, 1056-1065.

8. Hestenes, D., Wells, M., and Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, **30**, 141-158.

9 . Libarkin, J. C. and Anderson, S. W. (2006). The geoscience concept inventory: Application of Rasch analysis to concept inventory development in higher education. In X. Liu and W. J. Boone (editors), *Applications of Rasch measurement in science education*. Maple Grove, MN: JAM Press. pp. 45-73.

10. Libarkin, J. C. and Anderson, S. W. (2007). Science concept inventory development in higher education: A mixed-methods approach in the geosciences. Unpublished manuscript downloaded from http://newton.bhsu.edu/eps/gci.html#F

11. Maloney, D. P., O'Kura, T. L., Hieggelke, C. J., and Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *Physics Education Research, American Journal of Physics* Supplement, 69, S12-S-23.

12. Michael, J. A. (1998). Students' misconceptions about perceived physiological responses. *Advances in Physiology Education*, 19, 90-98.

13. Michael, J. A., Richardson, D., Rovick, A., Modell, H., Brucw, D. Horwitz, B., Hudson, M., Silverthorn, D., Whitescarver, S., and Williams, S. (1999). Undergraduate students' misconceptions about respiratory physiology. *Advances in Physiology Education*, **22**, S127-S135.

14. Michael, J. A., Wenderroth, M. P., Modell, H. I., Cliff, W., Horwitz, B., McHale, P., Richardson, D., Silverthorn, D., Williams, S., and Whitescarver, S. (2002). Undergraduates' understanding of cardiovascular phenomena. *Advances in Physiology Education*, **26**, 72-84.

15. Michael, J., Modell, H., McFarland, J., and Cliff, W. (2007). The "big ideas" of physiology. Part 1: What should students understand? This volume.

16. Pavelich, M., Jenkins, B., Birk, J., Bauer, R., and Krause, S. (2004). Development of a chemistry concept inventory for use in chemistry, materials and other engineering courses. *Proceedings of the 2004 American Society for Engineering Education Annual Conference & Exposition*. Paper #2004-1907.

FRAMEWORKS FOR REASONING AND ASSESSMENT IN MENDELIAN GENETICS

Joyce Parker, Charles Anderson, Merle Heidemann, Mark Urban-Lurain, John Merrill, Brett Merritt, Gail Richmond, and Duncan Sibley
Michigan State University

Step One – Defining Understanding

The first step in developing assessment instruments is to define the nature of the understanding to be assessed. The definition would serve three purposes:

- delineate desired outcomes for instructor and learners;
- guide analysis of learners' ideas; and
- provide a framework for instruction and learning.

When we began our work in 2003, we defined the understanding to be assessed in three ways – being able to:

- *define a concept and identify where it does and does not apply*;
- *use and move between standard representations of a scientific concept;* and
- *use a concept to explain multiple phenomena.*

The first definition of understanding grows out of the classical theory of concepts where concepts are characterized by their necessary and sufficient conditions (Bridgeman, 1927; Carnap, 1955; Jackendoff, 1989; Landau, 2000). The second definition, the prototype theory of concepts, is based on typicality (Smyth et al., 1994; Osherson and Smith, 1981; Lakoff, 1987). Prototypical representations in science include standard diagrams, models, analogies, or mathematical equations. For example, $6CO_2 + 6H_2O \rightarrow C_6H_{12}O_6 + 6O_2$ is a standard way of representing photosynthesis. The final definition of understanding focuses on utility (e.g. Lawson, 2000). What can the learner explain? This grows out of the knowledge-based model of concepts which likens concepts to scientific hypotheses subject to empirical verification or falsification (Wisniewski & Medin, 1994; Murphy & Medin, 1999; Johnson & Keil, 2000; Murphy, 2000).

Our early work showed that our initial definition of understanding served well the first purpose of defining desired outcomes. It was adequate when we used it to guide analysis of learners' ideas. We wrote questions that asked students to work with standard representations or apply a concept to a novel example. (Questions that asked students to define a concept were already readily available.) This led to a list of concepts and associated areas of misunderstanding. For example, for the idea of cellular respiration, students could define the concept and work with many of the standard representations, but very few could use the idea to explain mass loss in organisms. However for Mendelian genetics, specifically meiosis, students had difficulty producing or working with the standard "butterfly" representation of a duplicated chromosome. However the problem was that this diagnosis was not particularly useful for framing instruction. Each problem identified appeared to need its own instructional intervention.

A closer look at students' writing revealed a pattern that crossed topics: many responses violated simple scientific principles. Students' explanations had matter disappearing or appearing out of nowhere or changing in ways that violated basic rules of chemistry.
Energy came and went without appropriate explanation. When looking at genetics-related responses, information in the form of nucleotide sequences was replicated and changed without regard to very basic rules. This led us to be more specific about the type of reasoning we wanted students to use when explaining phenomena or using standard representations. We use the term ***principled reasoning*** to describe a type of understanding where students:

- *provide explanations that are commensurate with a few basic organizing principles as well as knowledge of biological events*.

In particular, we have found that tracing matter, energy, and/or information are useful organizing principles for cellular biology and genetics. By tracing matter we mean identifying the matter that is changing and the chemical identity of the molecules involved, describing the nature of the change, and conserving mass or accounting for mass changes. Tracing energy is a related principle that includes identifying the energy that is transformed and/o transferred and describing the nature of the change while conserving energy. Tracing information is a somewhat less familiar principle which we define as including identifying the nature of the information (nucleic acid or amino acid sequences, gradients, protein conformations, etc.) and how one type of information is translated into a specific signal. For example, the concentration of ATP is an indicator of the energy needs of a cell. High levels of ATP inhibit the allosteric enzyme phosphofructokinase, thus the information in the ATP concentration is translated or transferred to a conformational and activity change in the enzyme.

A Framework for Reasoning

The key ideas for principled reasoning about genetics are outlined in Figure 1. In this framework, the organizing principles of tracing matter and information are columns. Tracing energy is not an important organizing principle for this topic. The rows are levels or system scales. We see biology as a nested set of systems of decreasing scale. The framework shown deals with the subcellular through the organismal levels thus encompassing the content of an entire basic genetics course. An ecosystem level can be added to address population genetics and evolution. Each level is subdivided into the relevant processes. The last column echoes back to the level, identifying more specifically where the processes occur.

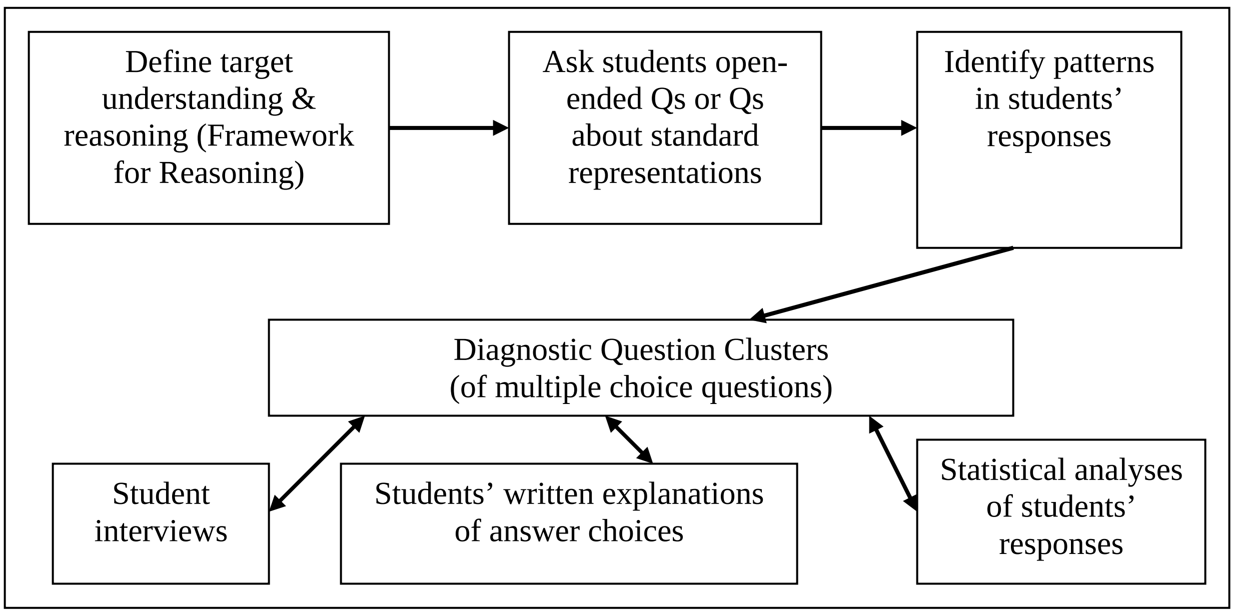
Figure 1. A framework for reasoning about genetics

	Tracing Matter	Tracing Information	Context / Location
ORGANISMAL LEVEL			
Phenotype	Interaction and regulation of **gene products,** beginning with fertilization, throughout the lifetime of an organism – observable or measurable traits.	Genetic information is variably expressed in time and space to "build" a functioning organism	All cells in all living organisms express particular traits
Gameto-genesis	Production of gametes by specific tissues/organs. Number of chromosomes is halved (see *meiosis*).	Genetic information: halved in particular cells.	Occurs in sex tissues/organs of variable complexity from plants, roundworms, flies, etc.
CELLULAR LEVEL			
Mitosis	One copy of each duplicated chromosome/DNA double helix distributed to daughter cells (see *replication*).	Duplicated genotype is distributed equally to daughter cells.	In the nucleus of eukaryotes; in the cytoplasm of prokaryotes.
Meiosis	Each chromosome pair* is halved independently. *in the case of ploidity, there will be more than two.	Genetic info: halved Alleles for each trait are separated following replication and pairing of homologous chromosomes, assuring one member of each pair in daughter cells	Occurs in the nucleus of eukaryotes.
Differential Gene Expression	Set of proteins/polypeptides translated in a cell.	Which ssDNA are templates for transcription is controlled by gene products, DNA modification, etc.	Eukaryotes - nucleus, prokaryotes - cytoplasm. Cytoplasmic interaction of gene products determines cell biochem.
Genotype	Set of nucleotide sequences that comprise the genes/alleles of each cell in an organism. Each allele/gene corresponds to a gene product.	Expression and interaction of allele products, polypeptides, determines trait expression (homo-, heterozygous, etc.)	In the nucleus of eukaryotes, in the cytoplasm in prokaryotes; expressed in cell primarily via translation
SUB-CELLULAR LEVEL			
Translation	Formation of a polypeptide chain, the gene product. Consists of a series of amino acids	Sets of three ribosomal nucleotides = anti-codons for amino acid sequence making up polypeptide chain	Occurs in the cytoplasm.
Transcription	Synthesis of mRNA from nucleotide subunits	ssDNA = information for mRNA synthesis (=allele/gene)	Occurs in the nucleus of eukaryotes, in the cytoplasm in prokaryotes.
Replication	Faithful doubling of DNA (or RNA) from pool of nucleotides in preparation for cell division.	ssDNA is template for formation of new double helix. Changes in nucleotide sequences of pre-gametic cells = mutation.	Occurs in the nucleus of eukaryotes, in the cytoplasm in prokaryotes.

Developing Questions that Assess the Defined Understanding

We use the frameworks for reasoning as ways of focusing and organizing the questions we develop. We aim for "diagnostic" question clusters, since our goal is to identify problems with students' principled reasoning. Figure 2 shows the general method we use to develop diagnostic question clusters. Once we defined the type of understanding and reasoning that we want to assess, we ask students open-ended application questions that require the use of the designated part of the framework and we look for patterns in their responses to these questions. Based on the patterns, we develop multiple choice items where the foils are intended to represent common problematic ways of thinking. We test the robustness of our explanation of why students choose particular foils by interviewing students, having them write about their answer choices, and by doing statistical analyses. Our goal is to develop clusters of multiple choice questions or at least easily gradable open-ended questions, because we deal with large enrollment courses.

Figure 2. Method for developing diagnostic question clusters



Current Work on Question Clusters for Mendelian Genetics and Meiosis

We have developed a question cluster for cellular respiration (Wilson et al, 2006; Wilson et al, in preparation) and are finishing a cluster on photosynthesis. Here we would like to showcase our current work on a cluster for Mendelian genetics, specifically the role of meiosis in explaining Mendel's law of independent segregation. In contrast to photosynthesis and respiration, we find that several of the standard representations are problematic for students. This is especially true when we ask students to keep track of both matter and information. Figure 3 shows an open ended question about meiosis and Figure 4 the patterns that we found among the students who failed to appropriately trace matter and/or information.

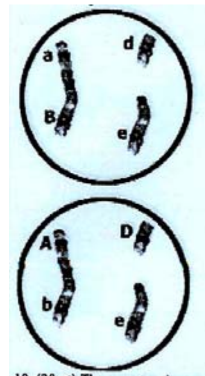
Figure 3. Open-ended question about meiosis

The gamete shown are TWO of the gametes produced when a single parent cell underwent a meiotic division.

A. Since meiosis of a single parent produces 4 gametes, draw a picture of the other Two gametes labeling the appropriate alleles.

B. Draw the parent cell in G1 phase prior to meiosis labeling homologous chromosome pairs and alleles as either dominant or recessive.

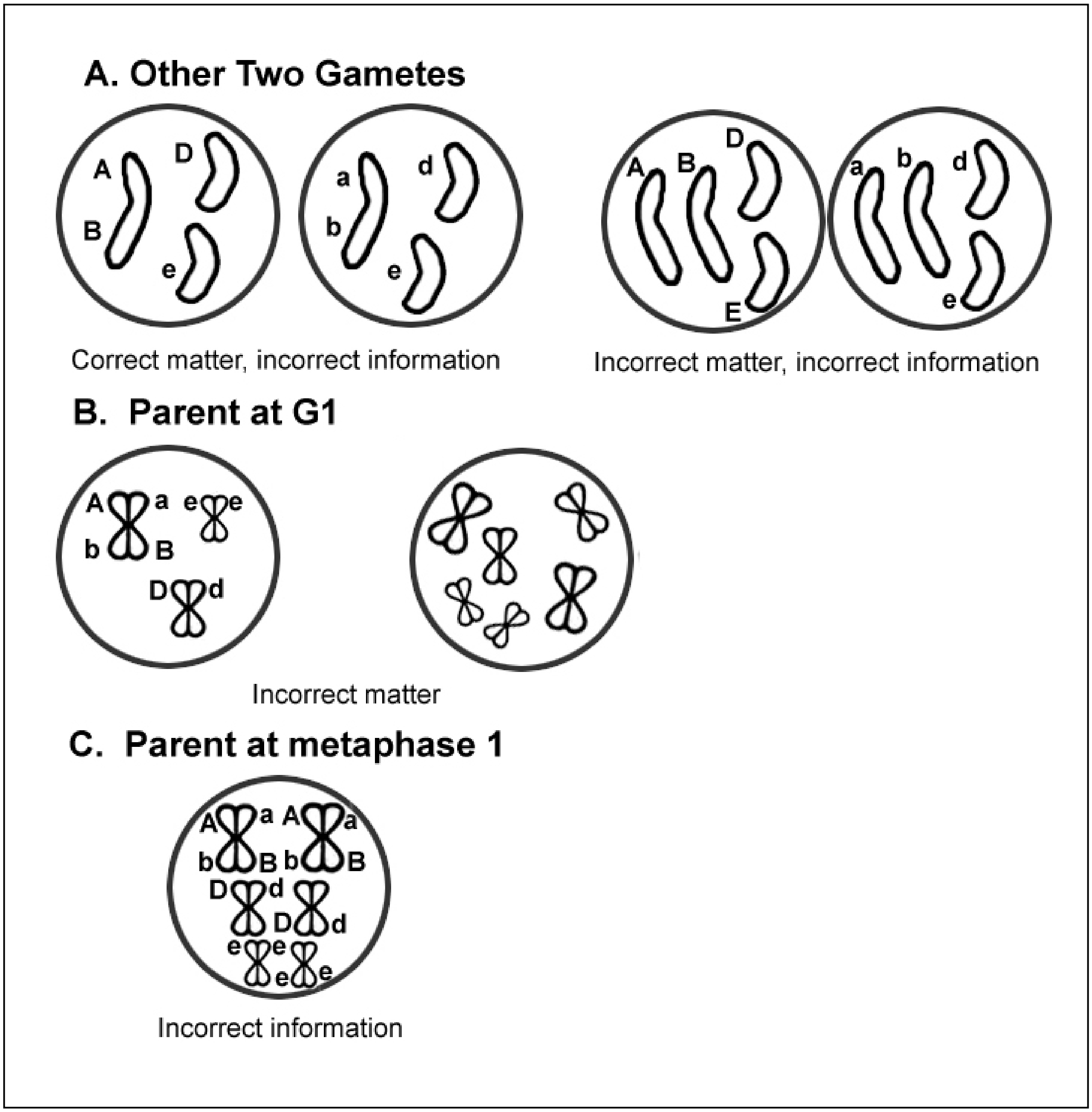
C. Draw the parent cell in metaphase I.



Students in an introductory biology class were given this question on an exam. The variation in their responses was huge. Without a framework it would be difficult to see patterns that could lead to the meaningful development of distractors. However our framework for reasoning suggests that categories such as correct matter (chromosome structure) – incorrect information (labeled alleles) might be useful. Figure 4 shows the most common categories of inaccurate responses to each part of the question. Pictures without designated alleles indicate that no one arrangement of alleles dominated. For part A, 51% of the 108 students drew the correct gametes, while 23% represented the matter correctly but had incorrect information/alleles, and the remaining 26% did not represent the matter correctly. For parts B and C, the examples given represent the most common responses, but not all of the responses.

We are currently drafting multiple choice versions of this question. The distractors for part A will include the responses shown here plus the other variations of incorrect alleles that resemble the first picture in Figure 4. The distractors for part B will include the first picture shown above as well as the many allelic variations of the second picture. Because the vast majority of students who answered part C incorrectly had the chromosome structure correct (paired dyads), all of the distractors will have this format with different variations of alleles. This cluster of questions will also include other multiple choice questions about gametogenesis and meiosis.

Figure 4. The most common incorrect responses to the question shown in Figure 3



Conclusions

We have been studying students' ideas about genetics for many years, but were originally unable to see patterns in their ideas that were useful for framing instruction. For example, it was clear that students needed more help interpreting the standard representations of chromosomes, particularly dyads. Students also confound double helices, homologs, and duplicated chromosomes. Many students who can use Punnett squares to solve inheritance problems cannot identify the possible gametes produced by the parents. These conceptual barriers that students encounter are related, but a unified instructional intervention was not evident until we looked across concepts and identified principled reasoning as a focus for assessment, a way of categorizing students' responses, and framing instruction.

Eliciting Big Ideas in Biology

David Niemi and Julia Phelan

For educators, researchers and others who are interested in improving learning, and particularly learning of complex skills and knowledge, a fundamental challenge is to understand how isolated skills and pieces of knowledge learned in a variety of classroom contexts can become inter-connected, meaningful, and generalizable. As cognitive science research has convincingly demonstrated, it is the connections among elements of knowledge, understanding of the meaning of important concepts and principles, and the ability to apply knowledge flexibly and effectively in a wide variety of situations, that characterize the development of knowledge toward greater expertise (e.g., Chi & Ceci, 1987; Chi, Glaser, & Rees, 1982; Glaser & Chi, 1988; Larkin, McDermott, Simon, & Simon, 1980; Mayer, 2005; Niemi, 1996; NRC, 2002, 2004). This is true for all domains that cognitive scientists have studied, including mathematics, science, history, reading, writing, other school subjects, and non-school domains such as race-track betting and chess. In fact these characteristics of expert knowledge--interconnectedness, understanding, and ability to transfer—are inextricably linked, a point that is critically important for educators and that constitutes an underlying theme in this paper.

The Importance of Big Ideas

Decades of cognitive research and educational experience have shown that when specific responses to specific tasks or questions are learned by rote, that knowledge does not generalize (e.g., Bassok & Holyoak 1989a, b; Bransford et al., 1999; Carpenter & Franke, 2001; Chi, Glaser, & Farr, 1988; Ericsson, 2002; Larkin, 1983; Newell, 1990;

NRC, 2004). Despite the robustness of this finding, however, science instruction in the U. S. has historically focused on the memorization of specific responses to specific questions, and as a result, the knowledge most students have is extremely context bound and not generalizable. Many students have not constructed the meaning of core concepts and principles and cannot relate concepts to problem solving skills and procedures.

In contrast to the piecemeal, context-bound knowledge that beginning learners have, expert knowledge has a relational structure: this is one of the strongest and most powerful conclusions to be drawn from decades of cognitive science research on the nature and development of knowledge—strongest, in the sense that it has been extensively and compellingly validated in a large number of studies, and powerful, in the sense that it has great explanatory force and broad implications for teaching, learning, and educational practice in general. For someone who has advanced knowledge in a domain, every element of that knowledge is connected to other elements in a highly organized structure, with certain statements, expressing important ideas, dominating and organizing other types of knowledge (e.g., Bereiter & Scardamalia, 1986; Chi & Ceci, 1987; Chi, Glaser, & Rees, 1982; Glaser & Chi, 1988; Larkin, McDermott, Simon, & Simon, 1980; Bransford et al., ; Niemi, 1996; Wineburg, 2002). That certain ideas organize other kinds of knowledge, including problem solving strategies and skills, was first and most dramatically revealed in a series of studies by Glaser and colleagues (Chi & Glaser, 1981; Chi et al., 1982).

In one study, for example, when physics experts and novices were asked to sort problems printed on index cards (Chi et al., 1981), the experts put together problems on the basis of abstract concepts and principles, e.g., Newton's laws, conservation of energy.

Novices, on the other hand, sorted on the basis of physical features of the problem situation, e.g., “there’s an inclined plane in these problems”. The novices either did not understand the theoretical principles or did not know how and when to apply them to problem solving situations. One effect of representing problems in terms of theoretical concepts is that expert problem solvers can activate and implement problem solving procedures linked to those concepts, e.g., formulas for solving conservation of energy problems. Novices have to resort to remembering how they solved problems with similar surface features, which can lead to ineffective solution strategies as problems with the same surface features (e.g., inclined planes) may be conceptually very different.

A follow-up study (Chi et al., 1982) using a different method, concept mapping, further confirmed that experts’ problem solving schemas were organized primarily around the laws of physics and conditions for applying them, while novice schemas were organized around the surface features of the problems. Chi et al. (1982) concluded that weakness in novices’ problem solving could be attributed primarily to deficiencies in their knowledge base and its organization. A range of studies have replicated these findings in other domains, and in each case researchers have found that experts have highly structured schemas, or knowledge structures, that are organized around central concepts or principles, or “big ideas”. The nature of these concepts differs from domain to domain, but in general they are abstract principles that can be used to organize broad areas of knowledge and make inferences in the domain, as well as determining strategies for solving a wide range of problems.

It has been known for many years that understanding of big ideas leads to more flexible and generalizable knowledge use, improves problem solving, makes it easier to make sense of and master new facts and procedures, and enables transfer (e.g., Ausubel, 1968; Chi & Ceci, 1987; Gelman & Lee Gattis, 1995; Larkin, McDermott, Simon, & Simon, 1980; Silver, 1981). The importance of understanding the core principles of a subject area and using them to organize knowledge (or for schema) has been demonstrated in many different subject areas, from interpreting X-rays, solving navigational problems, or playing chess (Chi, Glaser, & Farr, 1988; Ericsson, 2002; Larkin, McDermott, Simon, & Simon, 1980a,b) to mathematics (Ball & Bass, 2001; Carpenter, Fennema, & Franke, 1996; Carpenter & Franke, 2001; Collis & Romberg, 1991; diSessa & Minstrell; 1998; Lane, 1993; Porter, Kirst, Osthoff, Smithson, & Schneider, 1993).

Identifying Big Ideas

Given the importance of big ideas—organizing concepts and principles--in the organization of expert knowledge and in expert problem solving, the obvious question is how to determine what the big ideas are in a domain. Chi and Glaser ingeniously chose tasks that could be classified in terms of big ideas, but they did not offer a comprehensive list of big ideas in physics. One could follow their lead and administer a lot of tasks to a lot of experts (biologists, for example). Presumably the experts would identify the big ideas represented by those tasks, assuming that the tasks actually reflected and required knowledge of the big ideas. Without knowing what the big ideas are, however, it would be difficult to insure that the tasks comprehensively covered the big ideas in question. There would be no way to guarantee that one had the right set of tasks, without identifying ahead of time the big ideas to be targeted by those tasks. To address this problem, we developed a procedure for eliciting big ideas directly from experts. We did this because we found that neither state standards nor existing curricula make clear what the big ideas in a domain are, nor how other elements of knowledge relate to the big ideas. We reasoned that once we had worked with experts to identify the big ideas for a domain, we would be able to use the big ideas to develop a structured framework of knowledge for each domain, and that framework could then used to build a content blueprint for developing and integrating curriculum, assessment, and instruction.

In this paper we describe how we used this methodology to identify big ideas in biology. In collaboration with experts in several fields of math and science, we have endeavored to use the models of subject area knowledge from studies of expertise in different domains to help guide us in creating lists of big ideas. We have worked closely with different teams to identify key knowledge in a subject area. This knowledge has taken the form of lists of big ideas used to organize knowledge in a domain. Our methods have varied slightly over time and the ideal process has been gradually refined as we have worked to create multiple subject area big ideas documents. There follows a brief description of our general procedure for elucidating big ideas— in this case we were working on the biology big ideas.

For the earth science meeting we recruited two high school teachers and two university biology professors. In order to have a representative sample we selected experts from a range of specialties to allow for different perspectives and broader coverage of all aspects of a domain. Our instinct was that we should have a broader range of experts (within the constraints of the people available to work with us) in order to make sure our big ideas list did not become too focused on one particular sub-area of the discipline. As our stated goal was to arrive at a list of big ideas with which we can organize a large domain of knowledge—breadth of expertise was critical.

To begin, we presented a brief overview of the meeting goals and made sure all participants understood the concepts of big ideas. A short slide presentation was given to the participants outlining some of the key research findings on big ideas, for example, the previously mentioned study by Chi, Feltovich & Glaser (1981) where the problem-solving abilities in experts vs. novices were compared was presented and discussed (see Figure 1).

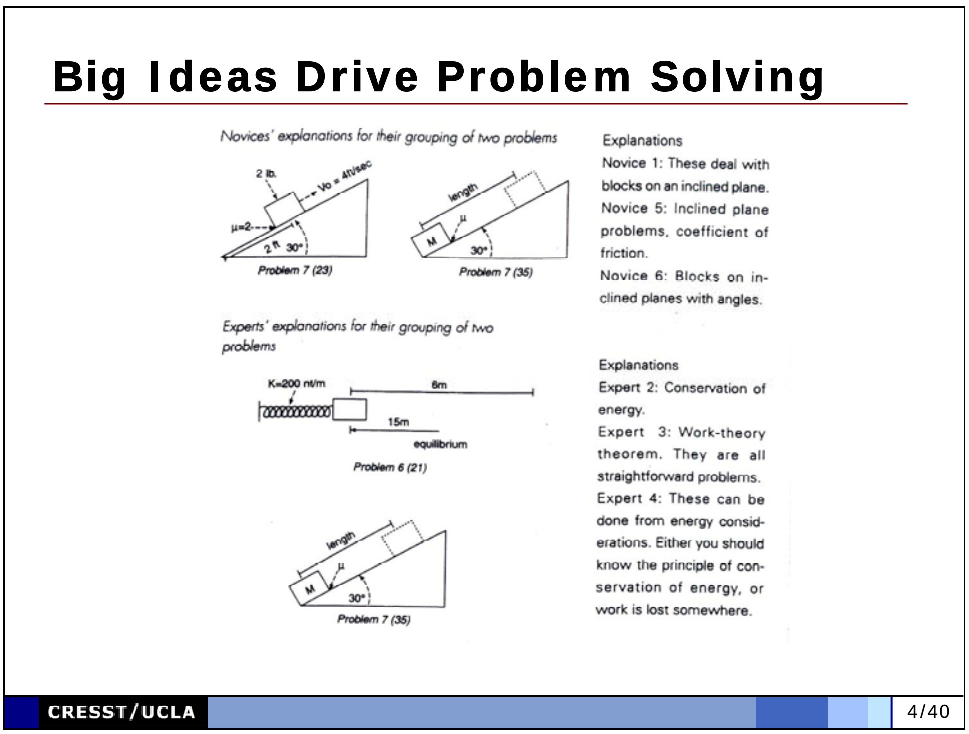


Figure 1: Example slide illustrating how experts and novices classify physics problems.

In this study, novices tended to group problems based on surface features—how the problems actually looked (in this case they classified two problems involving inclined planes together). Experts, on the other hand, grouped problems bases on the underlying theories and ideas needed to solve them (two conservation of energy problems for example). Big ideas were described to participants as concepts or principles used to organize thinking and other activities in a subject area. We emphasized the point that someone who understands the big ideas in a subject area can use them to make sense of and bring together facts that would otherwise be isolated, to identify and solve novel problems, to understand new information, and to create new knowledge.

In contrast, someone who has only memorized facts would not be able to accomplish any of these things, as cognitive science researchers have repeatedly shown. Rote facts can only be used to respond to questions that one has seen before, which is not a hallmark of advanced knowledge. We made clear to participants that we were not looking for a list of facts, or statements that might resemble a standards document, or a blueprint. Instead we were interested in how the participants in the meeting organized their thoughts and ideas about the subject area. Rather than consider how the ideas might be organized for pedagogical purposes, or how ideas are conventionally organized in curricula, the experts were asked to consider which ideas were most important in their own thinking. We provided examples of statements classified already as big ideas and those that would not be considered so (see Figure 2).

Big idea: Living things meet challenges of getting and using energy, reproducing, and maintaining their structure.

Something that would *not* be considered a big idea: Function.

Figure 2: Example Big Idea

The word *function* on its own would not be considered a big idea as it is not clear what about function is the important idea or concept. A better idea, if one wanted to make a statement about functions would be: *A function is a mapping between inputs and outputs such that each input is mapped to one and only one output*.

Following the introduction of the project and the initial presentation, we began a "brainstorming" session. The group was asked to propose some organizing principles within their domain of expertise. These were typed into a computer attached to a projector so all participants could see what was being recorded. This was critical as statements often needed to refined and edited several times before all members of the group were satisfied with them. It was also useful to allow participants to quickly state the ideas they thought were important—and then the language and exact wording could be changed as we went along. A critical component of this process was that these experts did not overly concern themselves with the ordering of concepts, or how they would be taught. Rather the emphasis was on the hierarchy of ideas around which one could design or organize courses of study at a later date. Several times during the meeting we could revisit this idea as participants tended to want to focus on what order you would teach concepts in.

The group of biologists very quickly came up with the theory of evolution as one of the overarching ideas in biology. Another big idea was the characteristics of life. When the preliminary list of big ideas had been determined, we, as a group, attempted to refine the statements and add in some supporting ideas. The supporting ideas are those that help further refine and explain the big ideas. An example of one of the big ideas and initial supporting content notes (in this case in earth science) is shown in Figure 3.

Big Idea: Plate tectonics is the theory that Earth’s surface is broken into pieces called plates that move and interact with each other in three basic ways: they collide, pull apart, and slide past each other. Plate tectonics provides the framework for understanding earthquakes, mountain building, volcanoes, and features of the ocean floor.

Supporting Ideas:

a) When earth’s surface collides it….

b) When earth’s surface pulls apart it…

c) Earthquakes occur when…

d) Volcanoes form when….

Figure 3: Big idea and supporting idea—initial notes.

Discussions took place over two days until a list of ideas was finalized. While there was some discussion on small issues (mostly on how to word, or define the supporting concepts), the experts agreed on the major organizing principles within the domain.

During the meeting, not all of the supporting ideas were fully developed. In most cases, some of the ideas were developed and others were left as rough headings. The experts created a list of headings they thought belonged with a certain big idea and this list was filled out following the session. As the supporting ideas were added, the list was sent around to the participating experts (those who had volunteered) for verification and modification as needed. Figure 4 is an example of a revised big idea and some of the supporting concepts for plate tectonics (the list of supporting ideas is an excerpt of a longer list):

Biological Use of Energy

Living organisms need energy to live. Some living things use sunlight for energy. Others get it from consuming other life forms. Energy from either source is not always directly usable. Living things convert some forms of energy into the chemical energy of those compounds that support life.

Supporting Ideas:

a) Cells perform three main kinds of functions that require energy: mechanical work, transport of molecules, and chemical transformations.

b) Chemical reactions that break bonds in organic molecules such as glucose can provide energy for chemical bonds in other molecules, such as adenosine triphosphate (ATP).

c) Adenosine triphosphate (ATP) is a complex organic molecule with energy stored in its triphosphate "tail." The energy stored in the bonds of these molecules of ATP is the most common direct source of energy for most cellular processes.

Figure 4: Example of big ideas and supporting idea generated by experts

To date big ideas meetings have been held in biology, physics, chemistry, geometry and algebra. These subsequent meetings were held in the same way as the first biology meeting, with the exception that for some subject areas we had a larger number of participants. At each meeting we found the breadth of experience and concentration within a discipline to be very beneficial. In all cases, we included both research-professors and teachers (at the college and high school levels) with different fields and study and areas of expertise. Experts agreed (sometimes after a period of discussion) on the major organizing principals in a domain, across all these subject areas. Much as the earth scientists agreed that plate tectonic theory was a very important big idea, the physicists tended to organize their thinking around the concepts of energy and matter. Similarly, life science experts agreed to on the importance of the theory of evolution and also the concepts of energy and how they relate to living things. Figure 5 shows examples of two big ideas, and related supporting ideas elucidated during the meetings with physicists and life scientists.

Physics: Motion: *Classical motion can be explained by applying Newton's laws of motion*

Supporting ideas:

a) Classical motion is the motion of everyday objects from atoms to galaxies.

b) Mechanics describes motion with velocity and acceleration

c) Acceleration is the rate of change of velocity and is a vector quantity

d) If the speed and/or direction of an object is changing, there will be some acceleration.

Life Science: **Chemical Basis for Life:** *Chemicals structured around carbon are the basis of life. The most important are carbohydrates, proteins, lipids and nucleic acids. Living things use these compounds in a water solution to meet challenges of getting and using energy, reproducing, and maintaining their structure.*

Supporting Ideas:

a) Organic compounds contain carbon. Carbon atoms are the building blocks of molecules. Their ability to combine in many ways with other carbon atoms and with other elements gives them the ability to make many molecules with a wide range of characteristics.

b) The chemistry of life takes place primarily in water solutions, and depends on the presence and properties of the water.

c) Carbon, hydrogen, oxygen, nitrogen, sulfur and phosphorus are the most common elements in organic compounds.

d) The four main classes of large molecules that make up living organisms are carbohydrates, proteins, lipids, and nucleic acids. Large molecules like these are called macromolecules.

Figure 5: Examples of big ideas and supporting statements

Instructional and Assessment Strategies associated with the big ideas

Unfortunately, school curricula, instruction and assessments do not always reflect what we know about subject area knowledge and how it develops. Too often, students are taught in a way that leads them to believe that learning means acquiring a huge number of meaningless facts and skills. As a result, most K-12 students never learn the big ideas in most subject areas (Schmidt, McKnight, & Raizen, 1997; U. S. Department of Education, 2002). Having carried out a rigorous analysis of domains to be taught, we have used them to many purposes including design of instructional materials and assessments. Most students need explicit instruction (e.g., clear definitions and explanations) on the big ideas in order to grasp them. And they need multiple opportunities to:

• Explain the big ideas and get feedback on their explanations

• Use the big ideas to understand and explain phenomena (e.g., how does the theory of plate tectonics account for earthquakes? what makes writing persuasive?)

• Identify situations in which big ideas apply and those in which they don't

• Figure out which facts and skills connect to which big ideas

The list of big ideas for a course can serve as an organizing framework for the course; e.g., as a way to review the objectives of the course to make sure that the big ideas are clearly and fully addressed. The big ideas themselves should be explicitly represented as course, unit, or lesson objectives in many cases it will require several objectives (not necessarily in the same unit) to adequately address a big idea.

The big ideas should also be present in the assessments. Asking a student to explain a situation using the concept of plate tectonics and perhaps apply it to a novel situation, would be allow the student to indicate a deeper level of understanding than would just asking for a definition, or having the student put some numbers into a formula. Figure 6 outlines the assessment strategies that can be used to assess understanding of the big ideas.

A. Define or state the big idea
B. Explain the meaning of each of the terms used to define or state the big idea
C. Explain the idea in their own words
D. Explain why the big idea is important
E. Recognize many different situations in which the big idea applies; e.g., situations that can be explained by the big idea
F. Apply the big idea to relevant situations and problems, including ones not encountered before
G. Use the big idea to solve problems and justify problem solving procedures (e.g., explain why a particular procedure works, using the big idea)
H. Recognize situations in which the big idea does not apply
I. Distinguish true from false statements about the big idea
J. Use the big idea to make inferences (e.g., infer causes) or predict consequences
K. Connect other concepts, facts, and skills with big ideas

Figure 6: Assessment strategies associated with the big ideas:

Further examples of how the big ideas can be used in both instruction and assessment will be addressed, alongside a more comprehensive look at the big ideas list in biology and how we can make a subject area map from these ideas.

References

Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Education Psychology, 51*, 267-272.

Ball, D. L., & Bass, H. (2001). What mathematical knowledge is entailed in teaching children to reason mathematically? In National Research Council, *Knowing and learning mathematics for teaching: Proceedings of a workshop* (pp. 26-34). Washington, DC: National Academy Press. Available: http://books.nap.edu/catalog/10050.html.

Bassok, M., & Holyoak, K. J. (1989a). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Memory, Learning, and Cognition*, *15*(1), 153-166.

Bassok, M., & Holyoak, K. J. (1989b). Transfer of domain-specific problem solving procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 522-533.

Bereiter, C., & Scardamalia, M. (1986). Educational relevance of the study of expertise. *Interchange, 17*(2), 10-19.

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal, 97*(1), 3-20.

Carpenter, T., & Franke, M. (2001). Developing algebraic reasoning in the elementary school. In H. Chick, K. Stacey, J. Vincent, & J. Vincent (Eds.), *Proceedings of the 12th ICMI Study Conference* (Vol. 1, pp. 155-162). Melbourne, Australia: The University of Melbourne.

Chi, M. T. H., & Ceci, S.J. (1987). Content knowledge: Its role, representation, and restructuring in memory development. *Advances in Child Development and Behavior*, *20*, 91-143.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7-75). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collis, K., & Romberg, T. A. (1991). Assessment of mathematical performance: An analysis of open-ended test items. In M. C. Wittrock & E. L. Baker (Eds.), *Testing and cognition* (pp. 82-130). Englewood Cliffs, NJ: Prentice Hall.

diSessa, A. & Minstrell, J. (1998). Cultivating conceptual change with benchmark lessons. In J. G. Greeno & S. Goldman (Eds.), *Thinking practices in learning and teaching science and mathematics* (pp. 155-187). Mahwah, NJ: Erlbaum.

Ericsson, K. A. (2002). Attaining excellence through deliberate practice: Insights from the study of expert performance. In M. Ferrari (Ed.), *The pursuit of excellence in education* (pp. 21-55). Hillsdale, NJ: Erlbaum.

Glaser, R., & Chi, M.T.H. (1988). Overview. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. xv-xxxvi). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Lane, S. (1993). The conceptual framework for the development of a mathematics performance assessment instrument. *Educational Measurement: Issues and Practice*, *12*(2), 16-23.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Models of competence in solving physics problems. *Cognitive Science*, *4*, 317-345.

Mayer, R. (2005).

National Research Council. (2002). *Learning and understanding. Improving advanced study of mathematics and science in U.S. high schools*. Committee on Programs for Advanced study of Mathematics and Science in American High Schools. J. P. Gollub, M. W. Bertenthal, J. B. Labov, & P. C. Curtis (Eds.). Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Niemi, D. (1996). Assessing conceptual understanding in mathematics: Representation, problem solutions, justifications, and explanations. *Journal of Educational Research*, *89*, 351-363.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Pellegrino, J. P., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment.* Washington, DC: National Academy Press.

Porter, A. C., Kirst, M. W., Osthoff, E. J., Smithson, J. S., & Schneider, S. A. (1993). *Reform up close: An analysis of high school mathematics and science classrooms.* (Final report to the National Science Foundation on Grant No. SPA-8953446 to the Consortium for Policy Research in Education.) Madison: University of Wisconsin-Madison, Wisconsin Center for Education Research.

Schmidt, W. H., McKnight, C. C. & Raizen, S. A. (1997). *A splintered vision: An investigation of U.S. science and mathematics education.* Boston: Kluwer Academic Publishers.

Silver, E. A. (1981). Recall of mathematical problem information: Solving related problems. *Journal for Research in Mathematics Education, 12*, 54-64.

Larkin, 1983; Newell, 1990; NRC, 2004).

An Assessment of Student Comprehension of the "Big Ideas" in Biology in the Context of General Education Courses[1]

Joelle C. Presson[1], Kathy McAdams[2], Pam Lanford[3], Therese Leslie[4], Dave Straney[5], Edgar Moctezuma[5], Janet Coffey[6] University of Maryland, College Park, MD

Abstract

This project involves the assessment of student comprehension of the "Big Ideas" in biology in a general education context, at the University of Maryland College Park. Assessment of general education at UMCP is governed by the learning goals set by faculty groups working at different levels of abstraction. First, faculty from departments across campus set learning goals that apply across general education categories. Next, science faculty established learning goals for the science general education category. Finally, faculty in the life sciences determined how the science learning goals should be assessed. This group developed an assessment of student comprehension of the "Big Ideas" in Biology, using a reading comprehension and writing format. Student comprehension will be assessed before and after taking a general education life sciences course. Pilot data indicate that this assessment approach can provide meaningful results, but that incentives for student participation are needed. The assessment will be completed fall 2007 and the results reported at the January CABII meeting.

Introduction

As at many other institutions of higher education across the U.S., the University of Maryland College Park has embraced assessment, both out of necessity from an accreditation perspective and because assessment will improve teaching and learning at UMCP. Assessment is happening at all curricular levels, in all degree and certificate granting programs. The guiding principle at UMCP is that assessments must be faculty-driven, and they must be meaningful to the faculty who design, oversee, and teach the various academic curricula. For example, in the College of Chemical and Life Sciences, the faculty first are assessing content knowledge in lower level courses, and quantitative skills at all levels. Critical thinking and research skills will be assessed in later years.

The assessment of general education learning goals is more problematic, for several reasons. The learning goals in general education are by nature more global and less tied to specific courses or sequences of courses than is true in major's curricula. In addition, the courses in general education categories often cut across departments and colleges, making cooperative faculty-driven assessments more cumbersome. A group of UMCP faculty and administrators spent a year thinking and talking about the problem, and came up with a general education assessment procedure that is consistent with our guiding assessment principles.

General education at UMP is currently organized into content areas, and so the assessment of general education is organized around the required disciplines. Students are required to take courses from a designated list in the areas of Humanities, Social Science and Social History, Math and Science, and Cultural Diversity. Each year the campus will assess one of these general education areas.

---

[1] College of Chemical and Life Sciences, [2]Division of Undergraduate Studies, [3]Department of Biology, [4]Department of Biology, [5]Department of Cell Biology and Molecular Genetics, [6]Department of Curriculum and Instruction.

To give the assessments context and to integrate the learning goals across the general education program, the first step was for learning goals to be set for the entire general education curriculum. Faculty groups met, discussed, and finalized these goals.

I. Overall general education learning goals

1. Demonstrate understanding of major findings and ideas in a variety of disciplines beyond the major;
2. Demonstrate understanding of methods, skills, tools and systems used in a variety of disciplines, and historical, theoretical, scientific, technological, philosophical, and ethical bases in a variety of disciplines;
3. Use appropriate technologies to conduct research on and communicate about topics and questions and to access, evaluate and manage information to prepare and present their work effectively to meet academic, personal, and professional needs;
4. Demonstrate critical analysis of arguments and evaluation of an argument's major assertions, its background assumptions, the evidence used to support its assertions, and its explanatory utility;
5. Understand and articulate the importance and influence of diversity within and among cultures and societies;
6. Understand and apply mathematical concepts and models; and
7. Communicate effectively, through written and oral communication and through other forms as appropriate.

Next faculty from the various science disciplines discussed and finalized the learning goals for the science general education category. These clearly build on the learning goals for the general education program overall.

1. Use quantitative information and/or mathematical analysis to obtain sound results and recognize questionable assumptions;
2. Demonstrate understanding of the broad principles of science and the ways scientists in a particular discipline conduct research;
3. Make observations, understand the fundamental elements of experiment design, generate and analyze data using appropriate quantitative tools, use abstract reasoning to interpret the data and formulae, and test hypotheses with scientific rigor;
4. Understand how findings and ideas in science can be applied to explain phenomena and events and influence the larger society;
5. Understand the role that human diversity plays in the practice and history of science;
6. Communicate about science using appropriate oral and written means; and
7. Demonstrate proficiency in the collection, interpretation, and presentation of scientific data.

The execution of assessment of these learning goals was also placed in the hands of faculty groups. The general model adopted parallels the approach being used to assess learning goals in the majors and certificate programs. Faculty who teach in the various general education disciplines are charged with designing and evaluating the assessments. Institutional support is provided to carryout the assessment. The faculty groups in each area are supported and guided by administrators familiar with assessment practices. One major guideline is that any given assessment project will be small in scope, much like any given scientific study is small in scope in order to obtain clear answers. Within this context a group of eight faculty, the authors of this paper, designed a specific assessment for the life sciences general education curriculum.

Methods

Despite the long list of learning goals, the major focus of most science general education courses is on science content. Not surprisingly, the faculty charged with designing the assessment in the area of life sciences general education gravitated toward assessment of mastery of content. This relates to learning goal #1 in the overall area and learning goal #4 in the physical and life sciences area. Discussion amongst this group of faculty quickly focused on the ability of students to recognize the "big ideas" in biology, to write about them in an articulate way, and to see their expression in popular scientific journalism.

The "big ides" in biology that this working group came up with and the instrument for assessing student understanding can best be described by simply sharing the instructions given to students. The articles referred to in these instructions were one-page summaries of a scientific paper obtained from a poplar science news publication, such as *Science News*. All of the students in one course read and responded to the same article.

<u>Read and React: A CORE Science Exercise</u>

1. Read the attached article.

2. Consider the following list of core concepts in science, and circle three of these concepts that are represented in the article.

Concept List

a. Knowledge about life is derived from the process of science.
b. Life emerges from chemistry.
c. Life is cellular.
d. Life forms adapt and evolve.
e. DNA directs development.
f. DNA is the basis for inheritance.
g. DNA is the basis for biological diversity.
h. DNA is responsible for unity and diversity of life.
i. There is an interaction between biology, culture, and environment.
j. Life requires energy.

3. Choose one concept you have circled. Write one paragraph in which you (a) fully explain the concept as you understand it, and (b) tell how this concept relates to the article. Be sure to cite specific examples of information in the article that relate to the concept you are writing about. Please write your reaction legibly on the back of this page.

4. Read over the paragraph you have written to be sure that you have mentioned specifics about the article for each of your selected concepts. Also check for clarity and accuracy in your writing.

5. Please provide the following information to help with data analysis. (All of your answers will be confidential, and results of this exercise will only be reported as aggregate data from large groups.)

a. Your major:__________________________________
   Please indicate whether this is your intended major ____ or your declared major ____.
b. The number of college credits you had at the beginning of this semester:___
c. Other science courses you have taken at the University of Maryland:

____

d. AP or undergraduate science courses you have taken at other schools:

e. Gender: female _______ male _________

Thank you for completing this exercise.

The assessment was designed to be done at the beginning of a life sciences general education course and at the end, to determine if taking the course had any impact on student understanding of these important concepts. Four general education life sciences courses were chosen as the targets for assessment:

ANTH220 Introduction to Biological Anthropology, a major's course also taken by many non-majors;
BSCI103 World of Biology, a strictly non-major's course;
BSCI105 Principle of Biology I, a major's course also taken by many non-majors;
BSCI124 Plant Biology, a strictly non-major's course.

In the original design it was intended to do the post-course assessment spring 2007 and do the pre-course assessment on a different group of students fall 2007. The spring 2007 assessment was carried out on the day of course evaluations in three of the courses, and on the day of the final in one course. Students were not given extra credit points for doing the assessment in three of the courses, and were given extra credit points in one course. Trained graduate students, using the rubric below, scored the writings. During the first grading session the graduate student graders discussed and modified the rubric until the category definitions were clear and they felt they could objectively apply the scoring criteria.

Rubric

	0 Inadequate comprehension	**1 Some comprehension**	**2 Basic comprehension**	**3 Good comprehension**
Ability to identify three relevant concepts	Student did not circle any relevant concept.	Student circled one relevant concept.	Student circled two relevant concepts.	Student circled three relevant concepts.
Ability to explain one key concept	Student did not explain the concept.	Student response contains elements of an accurate explanation of concept.	Student adequately explained concept.	Student fully explained the concept and included elaboration.
Ability to apply understanding of one key concept	Student made no relevant connections to the article.	The student made one or more connections to the article, but attempts at meaningful connections are inadequate.	The student began to discuss the concept in the context of the article, and included some supporting details.	The student response demonstrated sophisticated use of concept in the context of the article, including supporting details.

Ability to communicate understanding of one key concept using good English skills.	The student response was inarticulate.	The student response is not clear and contains many grammatical errors.	The student response is mostly clear and straightforward and contains some grammatical errors.	The student response is clear and straightforward and contains only few grammatical errors.

The results for spring were fully analyzed, as discussed below. However, it seemed evident from the quality of answers that many students did not put effort into the assessment. For this reason the entire project is being re-done this fall. The same courses, articles, and instructions are being used as for the pilot in the spring. Three major changes are being made, however. First, nearly all students in the selected courses will be sampled by asking students to take the assessment in their laboratory sections. Second, the pre-course assessment will be carried out on half of the students in the first weeks of the semester, and the post-course assessment will be carried out in the last week of the semester. Third, the students in each course will be given a small number of extra credit points for completing the assessment. The writings from the pre-course and post-course will be coded with numbers corresponding to each, and scored using the same rubric and by the same trained graduate students who scored the pilot writing samples. In this way the scorers will be blind to whether a given paper is pre-course or post-course, and the two will be scored in the same way.

Preliminary Results

Pilot data from the end of spring 2007 life sciences general education courses were obtained from the responses of 480 students. Students had no difficulty in identifying the biology concepts reflected in the article they read. Eighty percent of respondents circled three concepts. In general the concepts circled were relevant to the article they read, and did not include less relevant ideas. For example, fewer than 10% of the concepts circles were “DNA is the basis for diversity” or “DNA is responsible for the unity of life”. This response rate is appropriate as none of the articles that the students read directly dealt with the unity and diversity of life.

The students wrote an average of 473 words in response to the queries asking to explain and apply the concepts, although some students wrote only a sentence or two. The distributions of the abilities of students to explain and apply the biology concepts are shown in Figures 1 and 2, respectively. Note that at the end of the general education life sciences course only about 20% of students had a good understanding of the concept and only about 10% could communicate how the concept was related to the topic of the article. Writing communication scored somewhat higher. In this assessment the graders tried to separate use of language and clear wording from the content of the article. In practice this was not always easy to but the higher scores shown in Figure 3 indicate that our students write at a higher level than they understand.

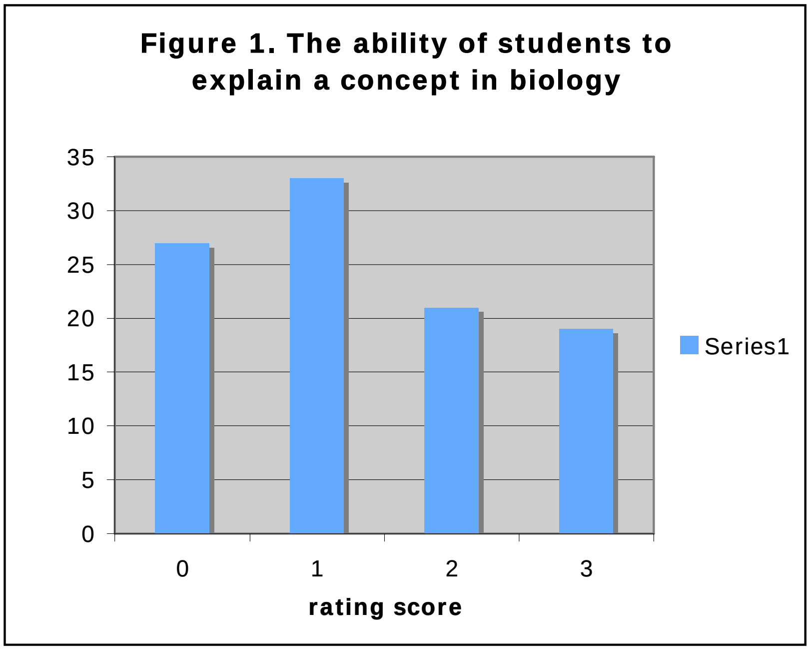


Figure 1. The ability of students to explain a concept in biology
35
30
25
20
15
10
5
0
0
1
2
3
rating score
Series1

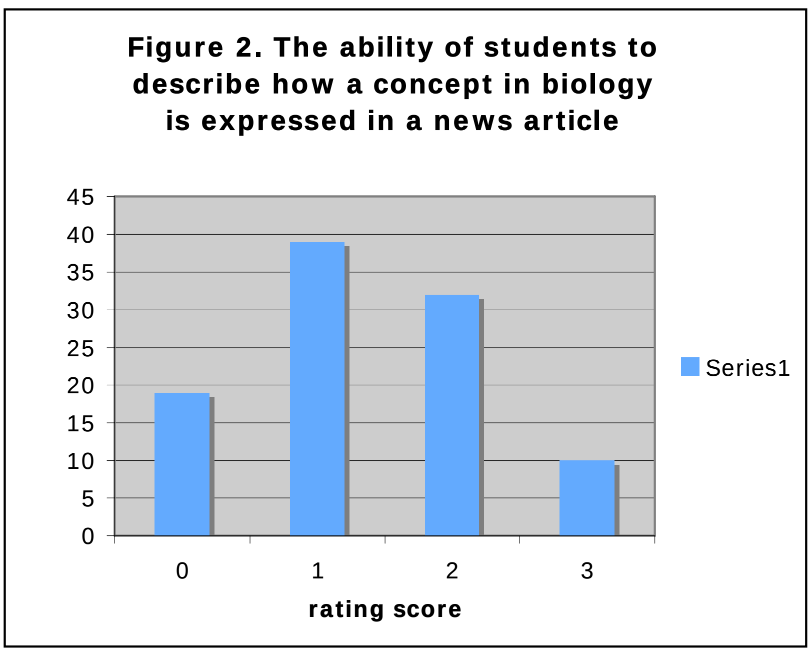
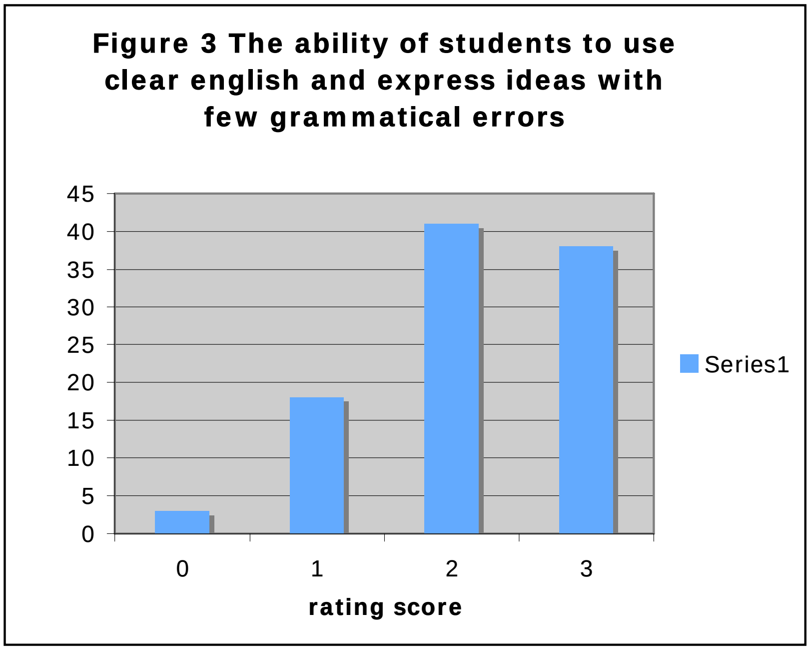


Figure 2. The ability of students to describe how a concept in biology is expressed in a news article
45
40
35
30
25
20
15
10
5
0
0
1
2
3
rating score
Series1



Discussion

One of the goals of any science general education curriculum is for students to understand, apply, and communicate the important conceptual ideas in the field. While it is difficult to say what the level of expectation should be regarding these abilities, the level of all three abilities in this pilot assessment are not where they should be. There are, however, some problems with this assessment that require a replication before the results are acted upon.

The assessments in three of the four courses were given on the day of course evaluations, and no points were given for participation. This presents two problems. First, many students do not attend the course evaluation and so this procedure does not sample all of the students. While it may not be necessary to sample all students, the assessment procedure we used is capable of handling all of the students registered in the four courses, and so it is desirable to do so if we can. Second, both the venue of the day of course evaluations and the lack of any points associated with the assessment undoubtedly reduced student motivation to take the assessment seriously and do their best. The six graduate student graders were in agreement that many of the papers seemed to reflect minimal effort. For these reasons the assessment is being completely redone this fall, as describe above. Results from this replication will be analyzed, and compared with the results from the spring 2007 pilot.

Connecting the "Big Ideas" in Biology with Those of Other Disciplines

Duane W. Sears
Department of Molecular, Cellular, and Developmental Biology
University of California
Santa Barbara, CA 93106-9610
sears@lifesci.ucsb.edu

In the extensive email dialog that ensued after the first CAB meeting in Boulder last March, a particularly intriguing exchange was recorded in May between Mike Klymkowski, Kathleen Fisher, and Joel Michael regarding the chemistry of bond-breaking and formation and whether such concepts are appropriate or even necessary concepts to include in the teaching of introductory biology. In his email on this particular matter, Mike raised some provocative points that I think are extremely relevant to the dialog that will (should) occur at the upcoming CAB II meeting. As to the "issues of bond-breaking and formation, and energy," Mike questioned whether (these concepts) are "biology," noting that if "chemistry did its job, we would not have to worry about it." Next he posed the provocative idea that it "may well be possible to teach introductory 'biological reaction dynamics' without even mentioning deltaG or equilibrium constants."

This was a shocking idea to someone like me who teaches biochemistry and who anticipates (even expects) that incoming students will be versed (if not well-versed) in basic chemical concepts such as these, especially considering that entering students have already completed required courses in introductory chemistry and biology as well as organic chemistry. Of course, I agree with Mike's point that "if chemistry did its job" we wouldn't have to worry about it. I also agree with his insinuation that chemistry doesn't do its job and that we do have to worry about this, based on my own experiences. Thus, it is my practice to give students "pre-instruction" assessment quizzes during the first week of classes just when they are starting out in our year-long General Biochemistry lecture series. In these assessments, they are asked to solve "chemically-related" problems based on fundamental chemical concepts. What we have found is that a surprising number of students lack a functional understanding of equilibrium reactions and their associated constants, and many have little or no understanding of how such reactions and constants relate to "biological reactions dynamics" (to quote Mike). Even after having the problems posted for a week, only <u>half</u> of the 208 students entering my Fall 2005 biochemistry course, for example, were able to arrive at a suitable quantitative answer for the following problem:

> "Calculate the pK_{dn} for a 0.01 M weak acid solution in pure water if conductivity measurements show that it is 1.16% dissociated at equilibrium." (Ans. pK_{dn} = 5.87 +/- 0.05)

Likewise, only a <u>third</u> of these students (again having the problem posted for a week) where able to find an adequate solution for the following problem:

> "Calculate the pH of a solution containing 1.0 M glycine and 0.5 M HCl with the pK_{dn} values for Gly being $pK_{dn}(\alpha\text{–}NH_3^+) = 9.8$ and $pK_{dn}(\alpha\text{–}COO^-) = 2.4$."
> (Ans. pH = 2.4 +/- 0.05)

I could easily describe many other related examples but the bottom line here is that we have noted similar difficulties with approximately the same percentages of students in other academic years (1). Thus, it seems that many students start out their junior or senior years in upper level biology courses with significant "chemical" conceptual difficulties. We have made some effort to identify the underlying causes, as discussed in a previous publication (2), but it is still clear to us that many biology students are challenged in terms of understanding the chemistry of <u>reversible</u> equilibrium reactions. I would argue that understanding fundamental "chemical" concepts, such as those needed to solve the problems above, is key to understanding important biological concepts in general. In fact, I would argue that the Gibbs free energy (deltaG) of <u>steady-state</u>, <u>irreversible</u> reactions is a "big idea" in biology because the related concepts (which took years for biochemists to fully appreciate) allow for thermodynamic explanations of how energy flows and drives most dynamic biological processes. The significance of zero free energy change for <u>reversible</u> equilibrium reactions and, hence, the underlying concept behind the equilibrium constant, are topics that typically are not covered well in introductory chemistry courses while the topic of steady state thermodynamics is usually approached only in biochemistry courses, if at all. However, "biological reaction dynamics" (again, to quote Mike) – help explain the ability of an organism or a cell to respond <u>reversibly</u> and usually immediately to chemical changes in its environment and also resist the inevitable pull of entropy that accompanies all natural processes. One can only understand these concepts in terms of the reversibility of non-equilibrium reactions. Thus, I would argue that such concepts embody "big ideas" in biology and need to be embedded in any biological concept inventory (BCI). I would also argue strongly <u>against</u> eliminating discussions of detaG and equilibrium constants in introductory biology courses and would recommend extending these concepts to include irreversible, steady reactions.

In order to construct and assess a meaningful BCI, we need to keep in mind that biology is really a complex amalgam of other scientific disciplines. While the highly successful Force Concept Inventory (FCI) developed by Hestenes and others (3) seems like a natural starting point and template for formulating a BCI, we must keep in mind that the FCI is much more narrowly defined in terms of concepts strictly derived from Newtonian mechanics, electricity, and magnetism. Biology's conceptual base is much broader by comparison. One cannot truly understand evolution without some knowledge of the geological history of the earth. One cannot understand evolution without knowledge of genetics. One cannot understand the mechanisms of genetics without knowledge of the "information carriers" of biology – DNA, RNA, and proteins. And, one cannot understand the chemical dynamics of biological processes without understanding the chemistry of reversible reactions, one form of chemical bond breaking and formation.

Thus, in facing the task of constructing a BCI, we face the challenge of adequately articulating the blend between concepts that are considered to be strictly in the domain of "biology" and those that draw from other fields of science. As a biochemist, I am highly interested in the development of a biochemistry concept inventory, but one that is firmly anchored in a broader Biology Concept Inventory. Ideally, the latter will logically underpin all of the major areas of biological specialization – biochemistry, genetics, cell biology, physiology, etc. In this way, the development of a BCI provides those of us who teach biology courses with a wonderful opportunity to add unprecedented coherence to the unwieldy and extremely vast body of knowledge we call biology. As Bill Bryson so neatly illustrates many times over in his “A Short History of Nearly Everything” (4), there is wonderful coherence in the natural world where trilobites, perhaps the most ancient of known forms of life, ultimately gave rise to australopithecines, among the earliest known hominid fossils, eventually leading to the appearance of modern man, whose entire genome has now been the decoded along with many other forms of life. With this kind of coherence, the BCI should connect living things and living processes in ways the help our students understand how the world was, how it is now as we know it, and how it likely will become given what we know now.

(1) Scott E. Thompson‡, Nathan J. Barrows S. Robin Saxon, and Duane W. Sears (2007) “Identifying Student Misconceptions in the Development of a Biochemistry Concept Inventory” (in preparation)
(2) D. W. Sears, S. E. Thompson and R. S. Saxon (2007) **“**Reversible Ligand Binding Reactions: Why Do Biochemistry Students Have Trouble Connecting the Dots?” *Bio. Mol. Biol. Ed. 35:*105-118.
(3) D. Hestenes, M. Wells, & G. Swackhamer (1992). Force Concept Inventory. The Physics Teacher” *30:*141–158.
(4) Bill Bryson (2003) “A Short History of Nearly Everything.” Broadway Books, NY.

A Faculty Team Works to Develop a Concept Inventory that Measures Understanding of Microbiology Relevant to Host Pathogen Interactions.

Ann C. Smith, Gili Marbach-Ad, Volker Briken, **Najib El-Sayed,** Kennneth Frauwirth, Brenda Frederickson, Steven Hutcheson, Lian-Yong Gao, Sam Joseph, Vincent Lee, Kevin S. McIver, David Mosser, B. Booth Quimby, Patricia Shields, Wenxia Song, Robert T. Yuan and Daniel C. Stein

Department of Cell Biology and Molecular Genetics, University of Maryland

Abstract

Our goal was to establish bridges among seven Host Pathogen Interaction (HPI) undergraduate courses in the Department of Cell Biology and Molecular Genetics allowing students to develop depth in the area of host pathogen interactions. Our group of faculty and graduate teaching assistants met monthly to focus on reviewing the curriculum of each class, to identify learning goals, choose common microbial systems to highlight, and create a system of shared resources. We designed an assessment tool that includes 18 multiple-choice questions with open-ended explanations. This tool was distributed via WebCT to 200 students in General Microbiology (introductory class) and 60 students in Bacterial Genetics (one of our HPI advanced classes). The student responses were collated and reviewed by our HPI faculty as a group. We met for one day (9am – 4:30pm) to score student responses for alternate conceptions and then to develop a Concept Inventory which includes 18 multiple-choice questions that use commonly held alternate conceptions as distractors. Through Fall 2006 and Spring 2007 we distributed the Concept Inventory before and following six of our HPI courses (we collected total of 477 surveys). The surveys were analyzed once again in a full day group meeting. The findings allowed us to evaluate students' meaningful learning and specifically evaluate our courses to examine if we are covering all the concepts that we believe are required for understanding HPI at a level of sophistication appropriate for microbiology majors. Since we are implementing innovations in our courses, the concept inventory will allow us to monitor the effects of our curriculum reform.

Introduction

This study involved the development of a diagnostic assessment tool (concept inventory) to measure level of understanding relative to host pathogen interactions after completion of a set of microbiology courses. At the University of Maryland, as a group of faculty with expertise and research programs in the area of Host Pathogen Interactions (HPI), we are responsible for teaching the undergraduate courses with HPI content (presently seven courses). In fall 2004, we formed a teaching group to bridge learning between our courses (Marbach-Ad et al., 2007). Our group includes: five full professors, three associate professors, five assistant professors, three instructors, and an assistant professor from the College of Education with expertise in science education, as well as several graduate students with a strong interest in teaching who have joined us for various projects.

Our goal was to create bridges which would eliminate excessive overlap in our offerings and support a model where concepts and ideas introduced in one course would become the foundation for concept development in successive courses. We worked on curriculum and teaching approaches simultaneously with assessment. In this paper we describe the development of our concept inventory, applicable to other programs interested in assessing learning in biology and microbiology, we highlight common alternative conceptions identified through development of our assessment tool and document our findings that were gathered through the HPI Concept Inventory distribution before and following six of our HPI courses.

Alternative conceptions are ideas that differ from the corresponding scientific explanations. They are usually held by a significant proportion of students and are highly resistant to instruction. At the same time, these alternative ideas can serve as anchoring conceptions (Clement, Brown, & Zietman, 1999; Redish, 2003) from which to move to a scientific conception when suitable instructional strategies are developed. An assessment tool that identifies alternative conceptions of students is desirable for teachers who are striving to promote constructivist learning in their classrooms (Mintzes,Wandersee, & Novak, 2000).

Whereas experienced teachers and professors are usually aware of students' conceptual difficulties, novice teachers are not. In addition, many teachers recognize the need to assess their students' naive understandings but do not do so because they lack the appropriate tools. Perhaps the most effective way to identify alternative conceptions is to ask students to respond to open-ended questions. Obviously, this is logistically impossible in large classes. Therefore, the goal of our group was to produce an assessment tool that would elicit information about student conceptions that paralleled information obtained in open-ended questions, but could be used efficiently with large classes. Our approach (see design/procedure section) was similar but not identical to the two-tier method advocated by Treagust (1988), Anderson et al. (2002) and Odom & Barrow (1995). This test style is attractive because it separates factual knowledge (Tier 1=facts) from reasons for choosing a particular fact (Tier 2=mechanisms and beliefs).

In this paper, we describe a realistic and comprehensive tool for measuring undergraduate college students' understanding of HPI concepts. We describe the development and evaluation of our assessment tool (a concept inventory) and include an example question from the current 18-item version of our HPI Concept Inventory. [We are happy to share the HPI Concept Inventory upon request]

The design/procedure

The model system for learning and courses involved

This study is part of a longitudinal project that was started in Fall 2004. The project involved seven HPI undergraduate courses. General microbiology serves as a prerequisite for all other six courses: Pathogenic Microbiology, Microbial Pathogenesis, Bacterial Genetics, Immunology, Immunology Lab and Epidemiology (see Table 1). Our teaching group met monthly with average attendance of thirteen members. It was decided that to help students build bridges between content presented in the various courses and to help students anchor new material with previously learned material, we would focus discussion of host-pathogen interactions in all courses on two organisms *(E. coli* and *Streptococcus* sp.). Further it was decided that each course should include methods that would expose/engage students to the scientific research process. Simultaneously with these goals we developed the HPI Concept Inventory that was intended to evaluate our progress.

Table 1: The seven undergraduate courses in the project focusing on various aspects of host-pathogen interactions

Course	Lecture	Laboratory	Discussion session	Annual Enrollment
General Microbiology* (BSCI 223)	+	+	On-line discussion	800
Microbial Genetics (BSCI 412)	+	+		60
Immunology (BSCI 422)	+		+	100
Immunology Laboratory (BSCI 423)		+		60
Epidemiology (BSCI 425)	+		+	70
Pathogenic Microbiology (BSCI 424)	+	+		100
Microbial Pathogenesis (BSCI 417)	+		Discussion based class	25

* The General Microbiology course serves as a prerequisite for all of the other upper level classes.

Constructing the concept inventory

The HPI Concept Inventory development involved three phases: (a) defining the content boundaries of the test, (b) obtaining information about students' alternative conceptions, and (c) developing the instrument.

(a) ***Defining the content boundaries of the test.*** We considered as a group this question: "What do we want our students to truly understand and remember 5 years after they have completed the set of ours courses?" We asked our selves this question to determine the "**big ideas**" for our project. Accordingly, we developed a list of 13 HPI concepts (Figure 1). We aimed at concepts that we believe are required for understanding HPI at a level of sophistication appropriate for microbiology majors. Content validity of the concepts was established by our complete HPI group.

(b) ***Obtaining information about students' alternative conceptions***. Based on the HPI concept list, a 23-item multiple-choice test with free response answers was developed. With the free open-ended response, we aimed to assess students' alternative conceptions, which later would be used as distractors in the final multiple-choice test. Therefore, each question had two-tiers. The first tier consisted of questions with two to five choices (there could be more than one correct answer); the second tier consisted of a requests for explanation (explain your answer, or defend your response). Each question covered one or more concepts from the HPI concept list (See Figure 2, two-tiered questions).

Figure 1: The 13 HPI concepts: the “big ideas” for our project.

1. The structural characteristics of a microbe are important in the pathogenicity of that microbe.
2. Diverse microbes use common themes to interact with the environment (host).
3. Microbes respond to forces of natural selection. Important responses include changes in virulence and antibiotic resistance.
4. Microbes adapt/respond to environment by altering gene expression.
5. Microbes have various strategies to cause disease.
6. Pathogens and host have evolved in a mutual fashion.
7. The cell wall and the cell membrane affect the bacterial response to the environment.
8. There is a distinction between a pathogen and a nonpathogen.
9. The environment will affect the phenotype (pathogenicity) of a bacterium.
10. Microbes adapt/respond to the environment by altering their metabolism.
11. Immune response has evolved to distinguish between self and nonself.
12. Immune response recognizes general properties (common themes vs. specific attributes: innate vs. adaptive).
13. Immune response memory is specific.

Figure 2: Example for two-tiered questions

1A. Selection of an antibiotic resistant organism is based upon a change in the
a. Phenotype
b. Genotype
c. Both
d. Neither
e. either
1B. Defend your response
Question 1 covers the HPI concepts: 3, 4, and 10.

2A. What determines a Gram stain reaction?
a. Distinction relating to bacterial structure
b. Distinction relating to bacterial function
c. both
2B. Defend your response.
Question 2 covers the HPI concept 1.

(c) ***Developing the final instrument***. The 23 questions were piloted with a small focus group of two graduate students and two undergraduate students. Results from this focus group were analyzed by the HPI teaching team. Our 23 questions were amended to 18 two-tier questions. In the Spring of 2006 the 18 question assessment was distributed via WebCT (course management system supported at UM) to 200 students in General Microbiology (introductory class) and 60 students in Bacterial Genetics (one of our HPI advanced classes). In order to limit the time requirement for the students in this pilot, the delivery of the questions was limited to 5 questions per student. For each question we received around 60 responses from the General Microbiology course, and 20 responses from the advanced course. The student responses were collated and reviewed by our HPI faculty as a group. We met for one day (9:00am – 4:30pm) to score student responses for alternative conceptions and then to develop multiple-choice questions that use commonly held alternative conceptions as distractors. The product of this day was an 18 multiple-choice concept inventory. This concept inventory was distributed in Fall 2006 and Spring 2007 before and following six of our HPI courses.

The Two-Tiered questionnaire data analysis

The two-tiered questionnaire that we developed was used to obtain information about students' alternative conceptions, and to develop the concept inventory to assess the success in teaching HPI concepts. Table 2 shows an example for analyzing the two-tier questions. For each question, we first counted the number of students selecting each choice in the multiple-choice part of the question (first tier). Note, that depending upon how students chose to defend their response, there could be more than one correct option among the multiple choices. This test wasn't intended to grade students, we were interested in finding what alternative conceptions students held, and sometimes the answer to the first tier was correct and the explanation was incorrect, and vice versa.

Table 2: An example for analysis of one two-tier question and response's frequencies

Question	**Number of student that choose this choice**	
1. Selection of an antibiotic resistant organism is based upon a change in the	General Micro	Bacterial Genetics
Phenotype	1	7
Genotype	38	4
Both	25	13
Neither	0	1
Either	4	0
Defend your response (response categories)	General Micro	Bacterial Genetics
Excellent response	21	16
Basic response, more required to indicate higher understanding	9	0
Students didn't understand that *selection* is based on phenotypes.	28	1
Student responses indicated that they did NOT understand that a change in phenotype is due to a change in genotype	3	6
Misconception was with the understanding of the differences between genotype and phenotype	9	1
Either student did not answer question or student response was COMPLETELY off the mark ("at sea")	3	0

In order to define categories for the second tier ("defend your response") responses we decided to use the technique of Hodder, Ebert-May, & Batzli (2006). We formed three small groups of three instructors. Each group received 5-6 questions to analyze. For each question the group read all the answers and established categories (level of correctness and alternative conceptions). Then, each member went through each response and categorized the response. Finally, the three members of the group compared their ratings and discussed responses to reach a consensus for each student response. Figure 3 shows two examples for common alternative conceptions for the question 1.

Figure 3: Example for students' alternative conceptions

1. Selection of an antibiotic resistant organism is based upon a change in the
   (a) Phenotype (b) Genotype (c) Both (d) Neither (e) either.

   1. ***Students didn't understand that <u>selection</u> is based on phenotypes***.
   One student that chooses *(b) Genotype* wrote: When an organism becomes resistant to antibiotics (when it acquires an antibiotic-resistant gene that has been inserted as a marker), the organism's genotype has been changed.

   2. ***Misconception was with the understanding of the differences between genotype and phenotype***.
   Student wrote: This must be a change in the genotype because having antibiotic resistance will not necessarily change the look of an organism (phenotype). It will merely allow it to survive in situations where the antibiotic is present.

Following the analysis of all questions, each group built two multiple-choice questions for the final assessment tool (The HPI Concept Inventory). These questions usually include the opening sentence or sentences of the previous question and four or five choices of response: one is the correct answer and three or four distractors that reflect the students' alternative conceptions revealed in the analysis. For example, one question developed from the information that is presented in Table 2 was:

Selection of antibiotic resistant transformed bacteria is based upon a change in the:

A. phenotype of the bacteria.
B. genotype of the bacteria.
C. phenotype and genotype of the bacteria.
D. genotype and physiology of the bacteria.
E. genotype and morphology of the bacteria.

The concept inventory data analysis and findings

In Fall 2006 and Spring 2007 we administered the 18 questions concept inventory to six of our courses. Students were offered extra credit points for completing the survey. We collected pre-post surveys from 477 students with the following distribution:

- BSCI223 (General Microbiology), Fall 2006: 127 students.
- BSCI424 (Pathogenic Microbiology), Fall 2006: 96 students.
- BSCI223 (General Microbiology), Spring 2007, 109 students.
- BSCI412 (Bacterial Genetics), Spring 2007, 45 students.
- BSCI422 (Immunology), Spring 2007, 48 students.
- BSCI425 (Epidemiology), Spring 2007, 52 students.

Student performance on Concept Inventory

Table 3 shows average scores for each course pre-post surveys (each correct question weigh 1 point, since we removed 2 questions from the analysis, see below explanation, the maximum points on the test is 16 points). Inspection of these data shows that in both semesters of BSCI 223(General microbiology pre-requisite course), the pre-post grades are similar. This is an important finding, since in the spring and the fall semesters we had different instructors in the course. For future analysis we can treat these courses as comparable courses. Encouragingly, using T-test analysis, we found that in four of our courses (both BSCI223 courses, BSCI424 and BSCI422) there was significant improvement on the scores of the concept inventory from pre test to post test. And students taking pre-test in the advanced courses retained the level of understanding gained in the pre-requisite course (scores on BSCI 223 post are around 7.0 and scores on all pre tests in advanced courses are around 7 or greater).

Table 3: Courses' average scores on the pre-post surveys

Pre-Post	223/ 06 N=127	425 N=52	424 N=96	422 N=48	412 N=45	223/ 07 N=109
Pre	4.9	6.6	7.3	9.2	7.8	4.7
Post	7.0***	6.6	8.7***	9.9*	7.6	7.3***

note: Values were calculated without data from questions 8 and 13.
Questions 8 and 13 had very low discrimination values (under .30 for most classes).
* P< 0.05 *** P< 0.001

Table 4 shows percentages of correct answers for each question in the pre-post surveys. Since, different members of our group teach different courses, we decided to look over the percentages of the correct answers for each question in each course, and examine two major aspects: 1. the ability of students that receive good grades on the overall test to answer correctly to a specific question (discrimination factor); 2. the concept/s that each question covers, and which courses cover this concept/s.

Review of questions that make up the Concept Inventory

To examine the quality and appropriateness of each question, we calculated the percentages of correct answers dividing the students in each course into three groups in terms of grades on the complete inventory. We grouped percentages into high (25%), medium, low

(25%). Then we calculated for each question the discrimination value from 0 to 1. If a question has a value below .30 it means that students who did well on the test (high performance group) performed poorly on this question. We reviewed all questions administered in each class and found that in every class two questions (8 and 13) provided poor discrimination (no one did well, see Table 4).

Table 4: Students' percentages of correct answer on pre-post concept inventory

Question	Pre-Post	223/ 06 N=127	425 N=52	424 N=96	422 N=48	412 N=45	223/ 07 N=109
1	Pre	25	36	32	56	35	12
	Post	30	42	24	67	44	26
2	Pre	9	15	25	37	20	3
	Post	24	17	46	35	29	21
3	Pre	18	31	33	48	49	21
	Post	27	38	40	50	60	33
4	Pre	87	85	87	92	87	77
	Post	88	90	88	87	84	87
5	Pre	7	23	24	50	31	10
	Post	29	31	26	65	33	28
6	Pre	60	65	82	79	80	65
	Post	78	69	85	77	71	83
7	Pre	28	40	42	56	38	13
	Post	42	31	57	52	33	29
8	Pre	17	23	23	12	13	16
	Post	17	6	25	12	16	32
9	Pre	23	48	59	50	47	25
	Post	46	36	63	56	49	56
10	Pre	41	58	71	75	73	40
	Post	61	58	78	69	73	78
11	Pre	18	42	45	48	56	23
	Post	39	50	59	60	53	38
12	Pre	62	42	45	52	62	64
	Post	49	35	65	50	51	72
13	Pre	11	21	17	23	27	9
	Post	12	15	26	21	9	11
14	Pre	28	33	41	48	47	37
	Post	35	33	52	52	38	43
15	Pre	10	31	30	69	29	12
	Post	40	36	38	75	27	20
16	Pre	19	23	36	48	33	17
	Post	34	21	48	60	24	32
17	Pre	43	60	57	79	60	38
	Post	65	50	60	81	58	57
18	Pre	13	29	24	46	36	17
	Post	16	25	40	58	31	20

Questions 8 and 13 were designed to address issues regarding bacterial metabolism (Figure 1, concept 10). The group reviewed and discussed the importance of concept 10 and analyzed the clarity and specific student alternate understandings addressed in both question 8 and 13. As a result of this discussion, question 8 was re-worded. Question 13 was left as is. It was decided that the question wording was clear. We feel that in fact even our best students do not understand this concept. This was truly an excellent question as it revealed to us a gap in our curriculum.

2. Alignment of curriculum with HPI concepts

To examine the concept coverage in our courses we decided to use the Curricular Alignment Matrix (Assessing Academic Programs in Higher Education, URL). Curriculum mapping makes it possible to identify where within the curriculum learning objectives are addressed. In other words, it provides a means to determine whether our concepts are *aligned* with the curriculum. We decided to use three level of coverage 1-3. In which

1. Concept was introduced. Goal was to provide introduction. Expectation is that student is aware of the concept but has not mastered a deep understanding.
2. Concept was "covered" in a manner to engage students. Expectation is that students understand and can apply this concept to a straightforward circumstance.
3. Concept was fully "covered" in a manner were students understanding was assessed. Expectation is that students have a deep understanding.

Each instructor reflected upon their own course with respect to the HPI concepts. Table 5 shows the Matrix from this data.

Table 5: The Curricular Alignment Matrix (**1** = Introduced, **2** = covered, **3** = fully covered)

Concept / Course (BSCI)	223	412	423	424	425	417	422
1. The structural characteristics of a microbe are important in the pathogenicity of that microbe.	1	2	-	1-3	-	3	2
2. Diverse microbes use common themes to interact with the environment (host).	1	1	-	1-2	-	3	0
3. Microbes respond to forces of natural selection. Important responses include changes in virulence and antibiotic resistance.	1	3	-	1-2	-	3	0
4. Microbes adapt/respond to environment by altering gene expression.	1	3	-	1-3	-	3	0
5. Microbes have various strategies to cause disease.	1	1	-	1-3	1	3	1
6. Pathogens and host have evolved in a mutual fashion.	1	1	1	1-2	1	3	1
7. The cell wall and the cell membrane affect the bacterial response to the environment.	1	2	-	1-3	-	3	0
8. There is a distinction between a pathogen and a nonpathogen.	1	1	3	1-3	-	3	2
9. The environment will affect the phenotype (pathogenicity) of a bacterium	1	1	-	1-3	-	3	0
10. Microbes adapt/respond to the environment by altering their metabolism.	1	1	-	-	-	1	0
11. Immune response has evolved to distinguish between self and nonself.	1	0	3	1	-	2	3
12. Immune response recognizes general properties (common themes vs. specific attributes: innate vs. adaptive).	1	0	3	1	-	3	3
13. Immune response memory is specific.	1	0	1	1	-	2	3

Discussion

The idea of a Concept Inventory as an assessment tool dates back to 1992, when the Force Concept Inventory (FCI) was developed to measure students' conceptual understanding of motion and force (Hestenes, Wells, & Swackhamer, 1992). The FCI has been used over the past decade by physicists at several institutions of higher learning to assess the effectiveness of different teaching methods. Similar multiple-choice exams have been developed in chemistry and biology (Anderson et al., 2002; Odom & Barrow, 1995). As a group of instructors who care about their teaching and have taken on the challenge to create a cohesive set of courses that result in meaningful learning of HPI concepts, we sought to assess the effectiveness of our teaching efforts. We believe that the process of constructing a concept inventory as a group has had great value not only in the production of the product, but in the conversation about teaching and learning within our group. Together we worked to articulate the most important concepts for undergraduates to grasp in order to develop meaningful learning of HPI (the "big ideas"). Together we reviewed students alternative conceptions revealed from the two tier test, and finally as a group we built a concept inventory. Through the first distribution of the concept inventory we already have some significant findings: Student scores indicate that overall understanding of concepts does increase within each course during a semester, and the level of understanding remains as student progress to more advanced courses. But even the highest scores are not what we would like of our graduates.

We believe that the concept inventory will work for us to monitor the progress that our students make as they complete each course (pre and post assessment) and as they move from course to course (comparison of post tests performances). The concept inventory from this year will serve as baseline data. We began work on the HPI inventory because our overall goal as a teaching group was to improve how we teach and ensure that our students learn the concepts that we considered essential for understanding of host pathogen interactions. As we began changing how we taught (changing curriculum and teaching methods) it was very clear we needed a tool to measure our progress. Although changes in courses has begun our most senior students in our most senior course achieved only an average score of 9/16 on the HPI inventory. Therefore we have a way to go! Work in progress includes adding to our courses active learning approaches, such as using clicker questions, case studies, team projects (Marbach-Ad et al., 2007). We consider the HPI inventory a tool that is essential in revealing success with each new teaching experiment.

What are our challenges?

- Relating information revealed from student responses to our curriculum and teaching methods.
- Detailed understanding of student performance: Is performance affected by gender, race, previous courses outside our 7 courses, overall GPA, major etc?

What help could we use from others?

- How do we encourage students to take the assessment seriously and perform at their best?

- How often can we use this same set of inventory questions? How would we determine if an alternate question is "equivalent"?
- How do we help students to understand the importance of assessment and counter "test fatigue"?
- Best delivery mechanism? We used WebCT, the campus has now moved to Blackboard. Are there delivery mechanisms that have statistic packages?
- What statistic diagnostics should we use?
- Does our survey cover "big ideas' of biology within the context of microbiology? Should we include some of the "big ideas" from biology surveys?

Reference

Assessing Academic Programs in Higher Education.
http://web.uconn.edu/assessment/mapping1.htm

Anderson, D. L., Fisher, K. M., Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. Journal of Research in Science Teaching, 39, 952-978.

Clement, J., Brown, D.E., & Zietman, A. (1989). Not all preconceptions are misconceptions: Finding ''anchoring conceptions'' for grounding instruction on students' intuitions. International Journal of Science Education, 11, 554–565.

Hodder, J., Ebert-May D., & Batzli J. (2006). Coding to analyze students' critical thinking. The Ecological Society of America, 162-163. www.frontiersinecology.org

Hestenes D., Wells M., & Swackhamer G. (1992). Force Concept Inventory. Physics Teacher, 30(3):141–158.

Marbach-Ad, G., Briken, V., Frauwirth, K., Gao, L., Hutcheson, S., Joseph, S., Mosser, D., Parent, B., Shields, P., Song, W., Stein, D., Swanson, K., Thompson, K., Yuan, R., & Smith A. (2007). A Faculty Team Works to Create Content Linkages among Various Courses to Increase Meaningful Learning of Targeted Concepts of Microbiology. CBE--Life Sciences Education, 6, 155–162.

Mintzes, J.L., Wandersee, J.H., & Novak, J.D.(Eds.) (2000). Assessing science understanding: A human constructivist view (pp. 198–223). San Diego: Academic Press.

Odom, A.L. & Barrow, L.H. (1995). The development and application of a two-tiered diagnostic test measuring college biology students' understanding of diffusion and osmosis following a course of instruction. Journal of Research in Science Teaching, 32, 45–61.

Redish E. F. (2003). Teaching Physics with the Physics Suite. John Wiley & Sons.

Treagust, D.F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. International Journal of Science Education, 10, 159–169.

Using diagnostic test items to assess conceptual understanding of basic biology ideas:

A plan for programmatic assessment

Kathy S. Williams, Kathleen Fisher, Dianne Anderson, Mike Smith

Introduction

University departments and programs are being encouraged to engage in systematic programmatic assessment, but little guidance is provided as to how to best achieve this goal. Assessing the effectiveness of an undergraduate program of study is different from assessing performance of individual students or faculty. What are the best indicators of success? What can provide the best insights into strategies for improving the program?

It is widely known that students come to the university with many naive conceptions about scientific topics, conceptions that are at odds with established scientific conceptions. Thus, one approach for evaluating science programs is to ask, “To what extent have the students, enrolled in a particular course of study, given up their naive preconceptions and committed to accepted scientific conceptions?” Science is often counter-intuitive and it can be difficult for an individual to let go of a strongly-held preconception and acquire a corresponding idea that is scientifically sound. Further, there is evidence that one idea is not readily replaced by the other, but rather that the two ideas (‘naïve’ and ‘sophisticated’) often exist side by side in an individual’s mental structure, each competing with the other (citation). One version may be recalled in the classroom, while the other is employed in other settings. Eventually, one of those competing ideas gains dominance and the other fades into disuse.

When the new idea gradually gains dominance, this extended process is known as conceptual change. On the other hand, sometimes a new idea is retained just as long as the class in which it was introduced, and then it fades away again. This is known as short-term or temporary learning. Research indicates that lecture teaching is a relatively ineffective means for producing conceptual change, whereas engaging students in hands-on problem-solving, especially with appropriate hands-on, minds-on experiences, can be much more effective. This is especially true when the activities are combined with peer collaboration and discussion.

Here we define “diagnostic tests” as multiple choice tests that have been developed to assess the extent to which students choose scientifically sound ideas over commonly held preconceptions. Diagnostic test items are often two-tiered, wherein the first stem or item asks about a scientific ‘fact’ and the second stem asks about the respondent’s reason for choosing a particular response option in the first item. The second item typically offers four response options, one of which is scientifically correct and three of which represent common alternative conceptions. Other diagnostic tests present an authentic scenario in a short paragraph and pose questions about it, also with each response option aligned with either the scientifically correct or alternative conceptions. Since the alternative conceptions are attractive to many test-takers, scores on diagnostic tests are considerably lower than scores on traditional, content-based tests. Student performance on diagnostic tests also changes slowly as a function of time in an academic program, as you will see below. The stability of test scores over time is one of the features that makes these tests attractive for programmatic assessment.

Since the instructional program in biology at our institution relies heavily on traditional lecture and lab formats, and since diagnostic tests provide a relatively stable measure of gains, and since diagnostic tests aim to assess deep understanding of important ideas (as opposed to more superficial knowledge of vocabulary), we feel that diagnostic tests can provide an effective measure of learning gains among our biology majors. Thus, we chose diagnostic testing as one indicator of programmatic assessment. Other indicators may be added subsequently. This paper provides an overview of our progress to date.

Description of Diagnostic Tests

If we wanted to study the process of conceptual change, we would need to examine that process as it occurs in individual students, observe the timing and nuances of the changes, and explore what specifically triggered the changes. But that is not our purpose. Instead, we are using measures of conceptual change to assess the effectiveness of our undergraduate biology program.

We are developing five diagnostic tests to assess understanding of: Natural Selection, Osmosis & Diffusion, Cell Division (processes of mitosis & meiosis), Energy & Matter Transformations, and Nature of Science. The first two tests are in final form, but the last three are still undergoing revision and refinement.

Natural Selection Diagnostic Test. Items were selected from the Conceptual Inventory of Natural Selection (‘CINS’, Anderson, Fisher & Norman, 2002), with minor modifications (with permission from the authors). Included are five questions about the Galapagos finches and five about the Venezuelan guppies. This test is now stable.

By ‘stability, we mean that the diagnostic test has been evaluated with several classes and meets our expectations that a) all incorrect responses are selected by some students, b) modest gains in proportion of students selecting the scientifically correct response are seen as students progress through the major, and c) the proportion of students selecting the correct response falls within the range of 30% to 70% (Kaplan & Saccuzzo, 1997). We are working on a fourth criterion at the present time, as described below, by conducting interviews to determine if students are interpreting the items and responses in the ways that are intended.

Osmosis & Diffusion Diagnostic Test. Items were selected from the Osmosis and Diffusion test (Odoms & Barrow 1995), with modifications (with permission from the authors). This diagnostic test consists of ten two-part questions, twenty questions in all. This test is now also stable.

Cell Division (mitosis & meiosis) Diagnostic Test. This test is being developed entirely by this team and is still under development, as we consider special issues related to this topic. A diagnostic test aims to identify students’ understanding of ideas as opposed to their memorization of biological terms. Thus, it is desirable to avoid using biological terminology or jargon in the test items. This is quite challenging in most areas of genetics including mitosis and meiosis. Consider the following item:

1. A human egg and sperm join together to form the beginnings of a brand new baby. The fertilized cell prepares to divide. Prior to the first division of the fertilized call, each maternal chromosome
    a. is copied precisely
    b. pairs with the corresponding paternal chromosome
    c. undergoes crossing over
    d. all of the above
    e. (a) and (b)

The first two sentences, '*A human egg and sperm join together to form the beginnings of a brand new baby. The fertilized cell prepares to divide,'* is intended to convey the fact that we are looking at cell division in a diploid somatic cell, and that would involve mitosis. But mention of *sperm* and *egg* seems to trigger thoughts of meiosis, and most students answer this item incorrectly. To prompt deeper thought about the nature of the cell, we may break the item into two parts as follows, creating a 3-part question series overall

1. A human egg and sperm join together to form the beginnings of a brand new baby. The resulting fertilized egg is a ...
   a. haploid somatic cell
   b. diploid somatic cell.
   c. haploid germ cell
   d. diploid germ cell.

2. As an embryo begins to form and its cells divide, each maternal chromosome
   a. is copied precisely
   b. pairs with the corresponding paternal chromosome
   c. undergoes crossing over
   d. all of the above
   e. (a) and (b)

3. The reason for my response is that
   a. maternal chromosomes typically pair with similar paternal chromosomes before a cell divides.
   b. crossing over assures a healthy degree of genetic variability.
   c. somatic (body) cell division generally reproduces each chromosome exactly.
   d. genetic variation is essential for species survival.
   e. all of these events occur in cell division.

This is the approach we have taken so far, focusing on the process of mitosis and meiosis. Our Biology colleagues now would like us to expand the array of diagnostic tests to address a larger view of cell division (including regulation), development, and molecular biology, and we are embarking on that next.

The *Energy & Matter Transformation Diagnostic Test* has been created on the basis of student short essay responses and published research on prevailing alternative conceptions (D. Ebert-May citation). Generally this test consists of tiered items, in which one item poses a question and the next item asks for the respondent's reasoning. We are planning student interviews to help complete the conception specification table and add additional question items.

The *Nature of Science Diagnostic Test* we are using now asks for simple Agree/Disagree responses. In all cases, items are designed to help us learn how our students are thinking about specific essential concepts. We are leaning toward developing a more focused *Nature of Biology* diagnostic test with more response options that are tied to more alternative conceptions. It is likely that we will need to interview students and ask them to answer open-ended questions to help with that too.

Description of Programs Being Assessed

The Biology Department offers a Biology B.S. with emphases in a) Cellular and Molecular Biology, b) Ecology, c) Evolution and Systematics, d) Marine Biology, and e) Zoology. Students may also earn a B.A. in Biology, or a B.S. in Microbiology. All programs are exactly the same in their lower division science course requirements, and all require the same 3 “core” upper division courses in ecology, evolution & genetics, and biochemistry, cell, & molecular biology. About half of all majors also take an upper division microbiology course as their required organismal biology course. We chose to evaluate knowledge gains in those 4 upper division courses.

Methods

As part of the validation process, the diagnostic tests are administered as pre- and post-course tests in both major and non-major biology courses at our large public university and at adjacent community colleges. Since classes at all participating schools are large, the students in each class are divided into four or five groups. Each group receives a different diagnostic pre-test, and then gets a post-test on the same topic.

Description of Diagnostic Test Development

The diagnostic test items focus on biology ideas that are either *basic* (e.g., osmosis & diffusion) or *big* (e.g., natural selection), meaning they are topics that serve as either foundational knowledge or over-arching frameworks that influence t learning of many ideas in biology. These topics are rarely being taught directly in the particular course in which the student is enrolled. Each test consists of about ten 2-part items or twenty 1-part items. As noted above, students are not graded on their performance. However, many instructors give students some fixed number of points as an incentive for taking the diagnostic tests.

We began by identifying known misconceptions (including both published findings and observations collected by the authors, especially via short answer essays). Next, the particular topics to be tested were identified. In two cases we began with existing diagnostic tests, as noted above. A specification table was created summarizing the topics of interest, common alternative conceptions, and the correct or scientific ideas (Table 1). We then worked to generate two items per topic. These are have been evaluated and refined, usually over the course of 2-3 semesters. These are our guidelines for test item creation:

1) In order to distinguish between deep level understanding of an idea versus memorization of the meaning of a particular term, we systematically avoid using biology jargon. To the best of our ability, questions are phrased in simple English. This is particularly challenging where genetics issues come into play, since there really are not commonly used terms for such ideas as gene, allele, and DNA.

2) The correct response should be unambiguous.

3) Each distracter (incorrect response) should reflect a common naïve conception and should be attractive to at least some students

Once a test is created, it is given to several expert reviewers to determine if all agree that there is one correct response. It is also given to a pool of students to determine if the items are clear and easily interpreted.

Description of Interviews [to be added]

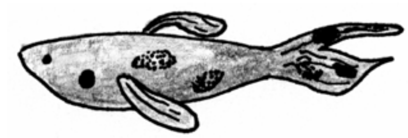
Table 1. Specification table for natural selection. Numbers in parentheses refer to corresponding item numbers.

Topic/Issue	Common Confusion or Naïve Conception (subtest item/stem #, option A-D)	Scientific Ideas (subtest item/stem #, option A-D)
A - Population stability	• All populations grow in size over time (1A, 7B) • Populations decrease (1D, 7C) • Populations always fluctuate widely/ randomly (1C, 7D)	• Most populations are normally stable in size except for seasonal fluctuations (1B, 7A)
B - Origin of Variation	• Mutations are adaptive responses to specific environmental agents (3C, 9D) • Mutations are intentional: an organism tries, needs, or wants to change genetically (3A, 3D, 9A, 9B)	• Random mutations and sexual reproduction produce variations; while many are harmful or of no consequence, a few are beneficial in some environments (3B, 9C)
C - Variation inheritable	• When a trait (organ) is no longer beneficial for survival, the offspring will not inherit the trait (4B) • Traits acquired during an organism's lifetime will be inherited by offspring (4A) • Traits that are positively influenced by the environment will be inherited by offspring (4D)	• Much variation is heritable (4C)
D - Origin of Species	• Organisms can intentionally become new species over time (an organism tries, wants, or needs to become a new species) (5C, 5D, 10A, 10C) • Speciation is a hypothetical idea (5B)	• An isolated population may change so much over time that it becomes a new species (5A, 10B)
E - Change in Population	• Changes in a population occur through a gradual change in all members of a population (2A, 8A) • Environment causes mutations to help individuals survive and reproduce (2D, 8C) • Mutations occur to meet the needs of the population (2C, 8D) • Learned behaviors are inherited (8C)	• The unequal ability of individuals to survive and reproduce will lead to gradual change in a population, with the proportion of individuals with favorable characteristics accumulating over the generations (2B, 8B)
F - Variation within a population	• All members of a population are nearly identical (6A) • Variations only affect outward appearance, don't influence survival (6B, 6C)	• Individuals of a population vary extensively in their characteristics (6D)

Any scientific data used in a test item is drawn from an actual scientific study. We feel this is important to keep our representations of scientific research 'authentic.' It is far too easy to misrepresent science when using hypothetical examples, and to thus risk introducing new misconceptions to students. This is illustrated in the following example about natural selection (Anderson et al. 2002), which briefly summarizes a study of guppies in South American streams.

Figure 1. Example of an item associated with a header describing a scientific study of guppies in South American streams. Several test items are linked to a single header.

Venezuelan Guppies



Guppies are small fish found in streams in Venezuela. Male guppies are brightly colored, with black, red, blue and iridescent (reflective) spots. Males cannot be too brightly colored or they will be seen and consumed by predators, but if they are too plain, females will choose other males. Natural selection and sexual selection push in opposite directions. When a guppy population lives in a stream in the absence of predators, the proportion of males that are bright and flashy increases in the population. If a few aggressive predators are added to the same stream, the proportion of bright-colored males decreases within about five months (3-4 generations). The effects of predators on guppy coloration have been studied in artificial ponds with mild, aggressive, and no predators, and by similar manipulations of natural stream environments (Endler, 1980).

Choose the one answer that best reflects how an evolutionary biologist would answer.

6. A typical natural population of guppies consists of hundreds of guppies. Which statement best describes the guppies of a single species in an isolated population?
   a. The guppies are identical to each other in all ways.
   b. The guppies share all of the essential characteristics of the species; the minor variations they possess don't affect survival or reproduction.
   c. The guppies are identical on the inside, but have many differences in appearance.
   d. The guppies share many essential characteristics, but also vary in many features. *

When two-tiered items are employed, the first tier (e.g., What happens when ___?) typically has two, and sometimes three, responses. In the second tier, the student is asked to give the reason for their response chosen in the first tier. Thus, the possible reason options\s given in the second tier should be equally divided between or applicable to the two initial response choices. An example from the osmosis and diffusion test is shown below.

Figure 2. Example of a two-tiered question in osmosis and diffusion.

5. If a small amount of salt (1 tsp) is added to a large container of water (1 gal or 2 liters) and allowed to set for several days without stirring, the salt molecules will
   a. be more concentrated at the bottom of the container.
   b. be evenly distributed throughout the container.
6. The reason for my answer is because
   a. there is movement of particles from a high to low concentration.
   b. the sugar is heavier than water and will sink.
   c. salt dissolves poorly or not at all in water.
   d. there will be more time for settling.

Mitosis and meiosis present special problems in trying to avoid use of jargon. Consider the example item below (Figure 3). We have avoided using the term 'somatic cell' or the more ambiguous term, 'body cell.' Instead we have talked about a cell in context. The reader has to interpret the events and figure out that this is a diploid cell now initiating the process of growth. Beginning with the terms *egg* and *sperm*, however, is likely to be misleading to the non-thoughtful reader. Given that a large proportion of upper division biology students seem to think that chromosomes pair in both mitosis and meiosis, there will be multiple reasons for incorrect responses.

Figure 3. An example item from the cell division diagnostic test.

1. A human egg and sperm join together to form the beginnings of a brand new baby. The fertilized cell prepares to divide. Prior to the first division of the fertilized call, each maternal chromosome
    a. is copied precisely
    b. pairs with the corresponding paternal chromosome
    c. undergoes crossing over
    d. all of the above
    e. (a) and (b)
2. The reason for my response is that
    a. maternal chromosomes typically pair with similar paternal chromosomes before a cell divides
    b. crossing over assures a healthy degree of genetic variability
    c. somatic (body) cell division generally reproduces each chromosome exactly
    d. genetic variation is essential for species survival
    e. maternal and paternal chromosomes are strongly attracted to each other

Diagnostic Test Administration

Once a test has been developed, it is administered in our "assessment" classes (lower division non-majors, and upper division majors core courses). Each item is then evaluated and refined as necessary. Any distracters that draw few or no responses are removed and replaced.

Assessments are administered as ungraded pre- and post-tests in all courses. In addition, we have some data from lower division non-major and major courses including general biology for biology majors and non-majors, collected at SDSU and local community colleges.

Since the relevant classes are large, we divide the students into subgroups, and each subgroup receives a different subtest. Thus, students are answering questions on topics not necessarily closely related to the topics covered the course. This is because the Biology Department is interested in learning about the basic capacities and knowledge of our biology students, in areas including biology, chemistry, math, and physics, to help us improve the overall biology curriculum, . Our goal is to analyze results using covariates indicating such things as which courses students previously completed, campus where lower division preparation was completed, and learning methods used in the prior courses as well as in their "current" course.

Initially, the diagnostic tests were deployed on paper and graded using ParScore, but soon after, we began using web-based tools. At first we used a university-based survey service, since our course management system (Blackboard) was not adequate at that time. Data were collected and evaluated for each test item using point bi-serial and difficulty values. The latest version of Blackboard now will produce a dataset of item responses that will let us conduct item analyses.

Evaluation and Analysis

Discriminability (point biserial values) and difficulty values are determined for all of the test items. The point biserial values indicate the ability of an individual item to discriminate between high and low performers on the entire test. The closer the point biserial value is to 1.00, the greater the discriminating power. Good test items generally result in point biserial values of between 0.30 and 0.70 (Kaplan & Saccuzzo, 1997). However, because the diagnostic tests are criterion-referenced test designed to identify concepts that students do or do not understand, rather than to discriminate among students, the point biserial values are of decreased usefulness (Gronlund, 1993).

The *difficulty* is determined by the proportion of students who respond to the item correctly.

The *reliability* of the test relates to the consistency of responses. A test must be shown to be reliable in order to be a valuable tool. As a measure of general internal consistency, we use the Kuder-Richardson 20. This method simultaneously considers all possible ways of splitting the test, so it improves on other methods of determining reliability in which the test is used only once (as opposed to test/retest methods). A good classroom test should have a reliability coefficient of 0.60 or higher (Gronlund, 1993).

Biology majors do not take their upper division courses in a specified order. For this reason, our current goal is to analyze results from each course using covariate analyses to determine the impact of such things as how many courses and which courses students previously completed, the campus where lower division preparation was completed, and the learning methods used in prior courses as well as in their "current" course.

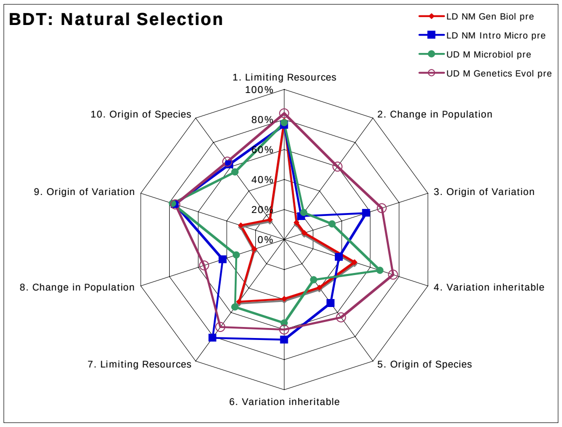
Results

Recent results demonstrate the striking consistency with which students are attracted to particular non-scientific conceptions (or "misconceptions"). Results from two of the subtests illustrate this.

Natural Selection Pretest Scores (Figure 3). The radar graph displays mean class scores on a 10-item test about natural selection, with zero in the center and 100% on the outer circle. The red (triangle) and blue (square) lines show scores for lower division biology classes, which occur in sequence. The green (closed circle), lavender (open circle) lines shows scores for upper division classes; students take these classes in any order.

Students in upper division Genetics and Evolution (Figure 3.) more accurately to questions 2, 3, 4, and 5, which ask about the origin and inheritance of traits and changes in traits over time (adaptation). Questions 2 and 8 are the most difficult for most students. They ask: what is the best way to characterize how a population changes over time, with #2 asking about finches and #8 about guppies (after Anderson et al. 2002). The questions illustrate the difficulties students have in recognizing that the *proportions* of organisms with different traits change in populations over time. Rather, they think that populations or individuals change.

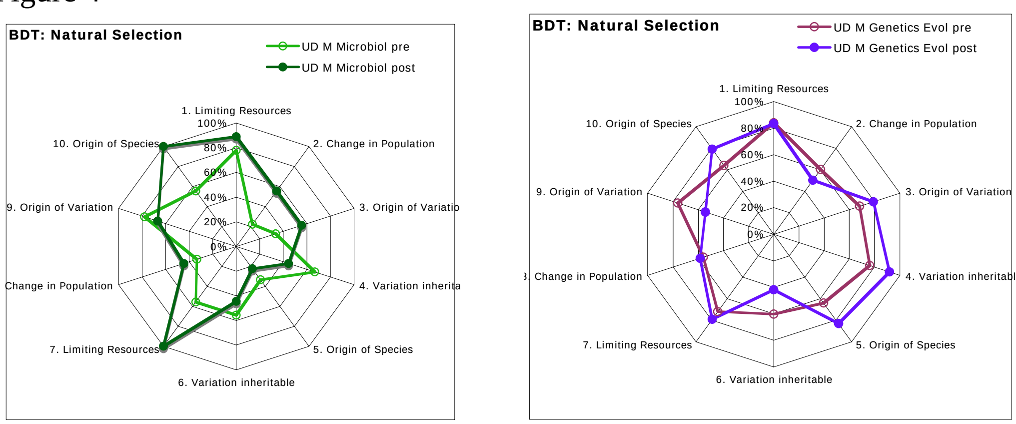
Figure 3.



In some classes we see changes occurring during the semester. Fig. 3 shows results from the upper division major's Genetics and Evolution course at the beginning and at the end of the course (pre and post). While students show the consistent difficulty with items 2 and 8, the pre and post scores are surprisingly high. We did see gains during the semester on some items (like 5 and 10 about origin of species) but the gains were not consistent on all items. This could be due to several issues. One is that there is no consistent sequence in which upper division students take these 3 courses. Since these results represent a set of students who voluntarily took this test, it is not a random sample, and we did not consider the course history of the students who took the survey

Natural Selection Pre- and Post-test Scores (Figure 4). Comparing pre- and post-test scores shows how the radar graphs can inform instructors about their learning in a semester. Students in Microbiology scored 100% on items 7 (carrying capacity) and 10 (origin of species), In contrast, Genetics and Evolution students showed gains on items 3 (origin of a trait), 4 (inheritance of genetic traits), and 5 (speciation).

Figure 4

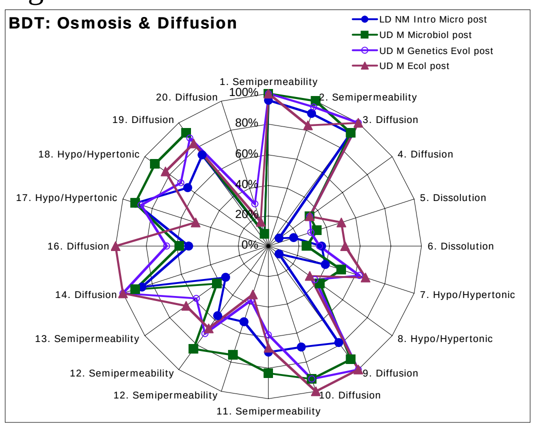


Because we did not account for which students were co-enrolled in other courses, it may be that students in Microbiology were concurrently enrolled in the Genetics and Evolution course. All participants were volunteers, not randomly sampled, so we cannot detect differences in incentives for participating in these results. However we do see some striking consistencies in some results (below).

Furthermore, following the constructivist idea, the outcome for instruction is still clear … explanations of the mechanisms of evolution, for example the source of variation and variation between members of a population, must give more treatment to why teleological and Lamarckian explanations are insufficient for explaining the unity and diversity of life. Students can then be guided to construct a more scientifically accurate understanding of the process of natural selection. We argue that with well developed diagnostic tests educators can identify "knowledge barriers" to address with effective instructional strategies in an appropriate curriculum sequence that will lead to increased knowledge gains on both diagnostic test items and more traditional items, such as on the Major Field Test (ETS). We hope to be able to make that comparison soon.

Osmosis and Diffusion Post-Course Scores (Figure 5). The osmosis and diffusion test has two-tiered items. Consider the ten diffusion questions. The first (odd-numbered) item asks a factual question, and the second (even-numbered) item asks for an explanation of why the response is correct. The power of the second items lies in the distracters that are chosen. Thus, students knew that molecules diffuse from high to low concentration in item 3, but many could not explain why this occurs in item 4. In items 5 and 6, students missed the fact that a small amount of salt will dissolve in a large container of water, as well as the reason why. They seem to know that heat speeds up diffusion and also why (9 & 10) and that blue dye will spread throughout the container and why (15 & 16). They know that concentration differentials affect diffusion (19), but much less sure about the reason why (20). There are clearly some concepts about diffusion that are not being learned or retained regardless of course. Others about semipermeability are apparently being learned by all. Comparing these results with those of a standardized test like the Major Field Test (ETS) will help confirm the acquisition or absence of understanding of those concepts.

Figure 3.



Discussion

We recognize that not all increases in mean class scores on diagnostic tests can be attributed to conceptual change in individual students. Some increases in average class performance will be produced by changes in the student population, as some students drop out of the program. We do not know which is the major contributor to increases in class scores, changing student populations or conceptual change in students. However, the desired end result of our undergraduate biology majors is to produce a group of students who have deep understandings of biology phenomena, consistent with what is known in the field of biology today. Both selective persistence in an undergraduate biology program and substantial intellectual effort on the part of each individual student are involved in producing this desired end result. Thus, we are using diagnostic tests to assess how successful we are in reaching this desired end point. We cannot claim that diagnostic tests are the best possible measurement tool for our purposes. What we can say is that we do not know of a better tool at this time.

Each student is tested at the beginning and end of each semester, with one form of one of the diagnostic tests. Thus, another possible concern is test fatigue and/or increasing test familiarity over time. This could contribute to increasing scores over time. On the other hand, students are often given points for completing a test, but their performance on a diagnostic test does not affect their grade. Thus, there is no particular motivation to either study for diagnostic tests or to cheat on them, or to try very hard to think about what the correct response may be. This could contribute to the lower scores seen on these tests as compared to traditional tests.

Diagnostic tests have, in come cases, revealed striking differences between classes of a single subject taught by different instructors and among different semesters. We are in the process of analyzing the background of those students, for example, to see if they took lower division preparatory science courses in smaller classes at community colleges or at SDSU.

Perhaps most importantly, diagnostic tests can give valuable feedback to individual instructors about what big biology ideas were learned successfully and which were not. Our challenge will be to generate interest among the faculty of our department to actually use the data we provide to modify their curricula and instruction. We are planning that approach and hope to report on those results soon.

Add more to discussion …

Relevant citations on diagnostic testing and assessing conceptual understanding in science

Anderson, D. 2003. Natural selection theory in non-majors biology: Instruction, assessment and conceptual difficulty. Dissertation for Ph.D. in Mathematics and Science Education submitted to San Diego State University and University of California – San Diego.

Anderson, D. L., 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.

Bloom, B.S. (Ed.). 1956. Taxonomy of Educational Objectives: The classification of educational goals. Handbook 1. Cognitive domain. New York: McKay.

Christianson, R. G. & Fisher, K. M. 1999. Comparison of student learning about diffusion and osmosis in constructivist and traditional classrooms. International Journal of Science Education 21 (6): 687-698.

Clement, J. 1982. Students' preconceptions in introductory mechanics. American Journal of Physics 50 (1): 66 - 71.

DeVellis, R. F. 1991. Scale Development. Sage Publications, Newbury Park, CA.

Driver, R., & Oldham, V. 1986. A constuctivist approach to curriculum development in science. Studies in Science Education 13: 105-122.

Fisher, K. M. & Lipson, J. I. 1986. Twenty questions about student errors. Journal of Research in Science Teaching 23 (9): 783-803.

Gronlund, N. E. 1993. How to Make Achievement Tests and Assessments. 5th Ed. Boston: Allyn and Bacon.

Hake, R.R. 1998. Interactive-engagement vs. traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses. American Journal of Physics 66: 64-74.

Hildebrand, A. C. 1989. Pictorial representations and understanding genetics: An expert-novice study of meiosis knowledge. Ph.D. Dissertation, University of California - Berkeley.

Horn, J. L. 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30: 179-185.

Johnson, B., & Christensen, L. 2000. Educational Research: Quantitative and Qualitative Approaches. Boston: Allyn & Bacon.

Kaplan, R. M. & Saccuzzo, D. P. 1997. Psychological Testing: Principles, applications, and issues. 4th Ed. Pacific Grove, CA: Brooks/Cole Publishing Company.

Lautenschlager, G. J. 1989. A comparison of alternatives to conducting Monte Carlo analyses for determining parallel analysis criteria. Mulitvariate Behavioral Research 24: 365-395.

National Research Council. 2000. Inquiry and the National Science Education Standards: A guide for teaching and learning (report). Washington, DC: National Academy Press.

National Research Council. 2003. Evaluating and Improving Undergraduate Teaching in Science, Technology, Engineering, and Mathematics. http://www.nap.edu/books/0309072778/html/

Novak, J. D., Mintzes, J. J., & Wandersee, J. H. 2000. Epilogue: On ways of assessing science understanding. In Mintzes, J. J. (Ed.), Assessing Science Understanding (pp. 355-374). New York: Academic Press.

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. Journal of Research in Science Teaching 32 (1): 45-61.

Palmer, D. H. 1999. Exploring the link between students' scientific and nonscientific conceptions. Science Education 83: 639-653.

Sadler, P. M. 1998. Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. Journal of Research in Science Teaching, 35: 265-296.

Sadler, P. M. 2000. The relevance of multiple-choice tests in assessing science understanding. In Mintzes, J. J. (Ed.), Assessing Science Understanding (pp. 249-278. New York: Academic Publishers.

Sundberg, M.D., and Dini, M.L. 1993. Science majors versus nonmajors: is there a difference? J. Coll. Sci. Teach. 23: 299 -304.

Sundberg, M D., Dini, M.L., & Li, E. 1994a. Decreasing course content improves student comprehension of science and attitudes towards science in freshman biology. Journal of Research in Science Teaching 31(6): 679-693.

Sundberg, M D. & Moncada, G.J.. 1994. Creating Effective Investigative Laboratories for Undergraduates. BioScience 44 (10): 698-704.

Sundberg, MD. 2002 Assessing student learning. Cell Biol Educ. 2002 Spring;1(1):11-5.

Sundberg, M.D. 2003. Strategies to Help Students Change Naive Alternative Conceptions about Evolution and Natural Selection. Reports of the National Center for Science Education 23(2): 23-26.

Tamir, P. 1971. An alternative approach to the construction of multiple choice test items. Journal of Biological Education, 5, 305-307.

Treagust, D. F. 1988. Development and use of diagnostic tests to evaluate students' misconceptions in science. International Journal of Science Education 10 (2): 159-169.

Students Learn About Their Own Bodies as Part of Their Biological and Citizenship Deduction. How do they learn? What do they learn first? From whom do they learn?

Subject

Students construct their ideas from formal or informal environments within an individual prospective or a social prospective (Scott, Asoko, and Leach, 2006). Scott, Asoko, and Leach (2006) state, “There are strong commonalities in how individuals appear to think about the natural world” (p. 38). The results from several studies in UK and US suggest students have commonalities about the order of organs of the digestive system and how the digestive system functions. In addition to commonalities, studies by Reiss and Tunnicliffe (2001, 2002) revealed that students’ understanding of the human organs and organ systems increases with age but remains incomplete.

When teachers or researchers ask subjects about their understandings of anything, subjects respond by presenting ‘representations.’ These representations may be words or mathematical symbols, drawings, physical constructions or even gestures or, in the language of Buckley *et al.* (1997) and Gilbert *et al.* (2000), expressed models – that is, representations of phenomena placed in a public domain. These expressed students’ models are presumed to be generated from mental models, i.e. the personal cognitive representations held by individual subjects. The only way for a researcher to understand a subject’s mental model of a particular phenomenon is by eliciting one or more of their expressed models of that phenomenon.

Procedure

1. Students answer questions posted on the Diagnoser website (www.Diagnoser.com). The questions are intended to elicit “facets of understanding” that students use to describe the respiratory system, circulatory system, and digestive system.

In the United States, the 751 students answered human biology questions posted on the Diagnoser website www.Diagnoser.com, developed by Jim Minstrell and his group Facet Innovation, Seattle, WA. The questions pertain to the respiratory system, circulatory system, and digestive system. Results presented are from the analysis of students’ answers to digestive system questions. The questions are intended to elicit “facets of understanding” that students use to describe the physical world. The program provides student feedback to either reinforce their thinking or encourage revision of their thinking. A class summary report is provided to teachers and a report on how each student answered the questions. The questions presented to students are intended to elicit the ideas (conceptions, misconceptions, facets of understanding) that the student uses to describe the physical world. Diagnoser is aligned with the National Science Education Science Standards (NAP, 1996) and Benchmarks (AAAS, 1993) for middle and high school students. Along with the questions, the program provides some feedback to students to either reinforce their thinking or encourage them to revise their thinking. The program also provides reports to teachers that show how each student answered the questions along with a class summary.

A*nalyses and Findings*

Results from Diagnoser showed students have alternative understandings of anatomy and physiology after instruction. Unfortunately, instructors take it for granted that all students would know the order of digestive organs after instruction. This is not true. Almost 50% of the students in grades 5, 7, 8, 9, and 10 either thought the large intestine came before the small intestine or the bladder came after the large intestine, or the lungs came between the esophagus and stomach. Fifty eight percent of students wrongly answered the question: most of the products of the digestive system are transported into the rest of the body in which part of the digestive system: (a) mouth 11%; (b) large intestine 20%; (c) stomach 24%; (d) small intestine 42%; (e) esophagus 3%.

Students had difficulty with abstract physiological concepts (theoretical conception) such as the function of micro-villi. Forty five percent of students answered the following question "In some diseases the tiny bumps (micro-villi) of the digestive system are destroyed. How would that affect the functioning of the digestive system?; (a) It would limit movement of stuff along the digestive tract (It would limit movement of stuff along the digestive tract, 45%; (b) It would limit absorption of stuff from the digestive tract, 51% (correct answer); (c) It would limit the mixing of stuff within the digestive tract, 4%."

Language can be confusing to students such as the "break down" of food, chemical (break-down of molecule) and mechanical (break-down of large pieces to smaller pieces, no change in molecular structure). The digestive system breaks the food you ate into smaller pieces of that food and the smaller pieces get transported into the body; (a) true 53%; (b) false 47%, indicates students are confused.

Conclusions

Students' exposure to information about human biology before beginning school can be from other sources of knowledge outside a formal learning environment. For example, Hannah (8yrs) describes asthma because her friend had kit and shows an understanding of respiratory system and mentions funny bone from personal experience. Several times students hear repeated phrases " if you didn't have bone you'd be jelly on the floor' blood made in thighs (influence of particular teacher year 4)."

It appears that there is the comfort level at which students' knowledge suffices and there is from anecdotal evidence the same leveling off of knowledge about phenomena and artifacts in many areas. A sufficiency but not a proficiency of knowledge is the norm.

Implications for Teachers

Instruction should start from learning organ and organ location but at the same time instruction should include a description about how the organ works within the organ system and it functions with all the systems, a holistic instructional approach. For example, the diaphragm is the muscle most often indicated on drawings, because it is taught with the respiratory system and the classic longitudinal section of the upper torso is the one studied and recalled, with the diaphragm as its lowered boundary. Diaphragm function provides the perfect opportunity for discussion of flow, specifically air flow, influenced by volume and pressure. Flow, volume, and pressure relationships are similar in other systems. Therefore, when students learn about the diaphragm they need to learn the functions of the diaphragm.

References

Abell, S.K., and Lederman, N. G. (Eds.) (2007) *Handbook of Research on Science Education.* London, UK: Lawrence Erlbaum Publishing.

American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for Science Literacy A Tool for Curriculum Reform*. New York: Oxford University Press.

Retrieved 2006 from: **http://www.project2061.org/publications/bsl/online.**

Buckley, B., Boulter, C. and Gilbert, J. (1997) Towards a typology of models for science education. In J. Gilbert (ed.) *Exploring Models and Modeling in Science and Technology Education* (Reading: University of Reading), 90-105.

Gilbert, J. K., Boulter, C. J. and Elmer, R. (2000) Positioning models in science education and in design and technology. In J. K. Gilbert and C. J. Boulter (eds) *Developing Models in Science Education* (Dordrecht: Kluwer), 3-17.

Minstrell, J. (2006) Facet Innovations. Diagnoser. www.diagnoser.com.

National Research Council (NRC). (1996). *The National Science Education Standards.* Washington, DC: National Academy Press.

Reiss, M. J., Tunnicliffe, S. D. (2001) What sorts of worlds do we live in nowadays? Teaching biology in a post-modern age. *Journal of Biological Education*. 35(3), 125-130.

Reiss, M. J., Tunnicliffe, S. D., Andersen, A. M., Bartoszeck, A., Carvalho, G. S., Chen, S., Jarman, R., Jonsson, S., Manokore, V., Marchenko, N., Mulemwa, J., Novikova, T., Otuka, J., Teppa, S., Van Rooy, W. (2002) An international study of young people's drawings of what is inside themselves. Journal of Biological Education. 36 (2), 58-65.

Tunnicliffe, S. D. and Reiss, M. J. (1999a). Students' understandings about animal skeletons. *International Journal of Science Education*, 21, 1187-1200.

Tunnicliffe, S. D. and Reiss, M. J. (1999b) Building a model of the environment: How do children see animals? *Journal of Biological Education*, 33, 142-148.

Tunnicliffe S.D. and Reiss. M.J. (2006) *Drawing breath: the use of drawings and interviews in a six-year longitudinal study of 5-11 year-olds' understandings of what is inside themselves.* Paper given at ERIDOB, September, Institute of Education, London.