Assessment of Model-Based Reasoning in Biology

CCLI-ASA-Area 1: New Development Proposal

Randall Phillis, Associate Prof. of Biology, University of Massachusetts, Amherst, MA
Neil Stillings, Prof. of Cognitive Science, Hampshire College, Amherst, MA

In recent years teachers and educational researchers have focused increasingly on a set of ambitious objectives for science learning, including the active mastery of concepts and the ability to engage in scientific inquiry (AAAS, 1993; Advisory Committee, 1996; National Research Council, 1996). Many questions remain, however, about how to make progress on these objectives in specific disciplines and under specific institutional constraints. In particular, the design and validation of appropriate assessments for higher-level learning goals is a continuing challenge (Pellegrino, Chudowsky, & Glaser, 2001). In this Area 1: New Development project we propose to develop, test, and disseminate methods for aligning formative and summative student assessment in large college biology courses with a particular primary learning objective, which we call Model-Based Reasoning in Biology (MBR).

Model-Based Reasoning as a Primary Learning Objective for College Biology

New approaches to assessment can be developed and tested effectively only in environments where new approaches to learning have been thoroughly implemented. We propose a program of assessment research and development in a mature instructional environment characterized by well-defined higher-order learning goals and classroom-tested learning activities. In this section we describe the learning environment in detail in order to clarify the assessment issues that are addressed in the proposed project.

The Setting for the Proposed Project

Over the past 5 years, with the support of the Pew Center for Academic Transformation and the NSF (Grant No. REC-9980519), we have introduced comprehensive reforms into the year-long introductory Biology sequence at the University of Massachusetts, Amherst. The non-laboratory class meetings of the two-semester introductory biology sequence at Massachusetts are held in lecture halls with 250-500 students. Under these circumstances it is difficult to avoid superficiality in the classroom and memory-oriented assessment. If more ambitious learning goals could be realized, approximately 800 students per year would be affected, and instructors for intermediate courses in a range of programs could adjust their learning goals upward. Given the high student/faculty ratio and the necessity of broad disciplinary coverage, however, the revised learning objectives for the course must be sharply focused. During our reform project it has emerged that a feasible higher-order cognitive objective for a large biology course is learning to reason with biological models (model-based reasoning in biology, or, MBR). We believe that our reforms are suitable for large biology courses generally. The thrust of the current proposal is the development, validation, and dissemination of reliable, valid, and practical methods for formative and summative assessment in MBR-oriented instruction.

Model-Based Reasoning in Biology

The explanation of natural phenomena in terms of underlying causal mechanisms is a central goal of modern science. The terms *theory, model*, and *hypothesis* are used somewhat interchangeably to denote explanatory constructs in science. We use the term *model* here to connote the conception of a relatively small set of interacting causal mechanisms that a scientist can reason with to generate predictions or interpret results in an area of research.

A revealing way to think about the concepts that are introduced in the introductory biology course is that they are packaged into theoretical models that capture the basic mechanisms involved in key biological processes. For example, the "central dogma" of molecular biology is a theory, or model, of the flow of genetic information from DNA to RNA to protein, comprising the *transcription* of DNA to RNA, and the *translation* of RNA to protein. Even this abstract statement of the model expresses highly significant generalizations about the biological processes within cells. In addition it serves to organize more-detailed models of the component processes. A simple model of *transcription*, for example, might specify the ordering of bases along strands of DNA and the construction of a strand of RNA through the pairing of complementary bases. And this model can in turn be elaborated to specify the process of transcription at various levels of detail.

For present purposes biological models, such as the model of gene expression sketched above, can be thought of as psychological entities that play a central role in scientific cognition. An expert biologist has a *mental* model of gene expression that is deployed in his or her scientific reasoning. An expert mental model is an integrated set of concepts concerning physical structures (molecules, membranes, cells, organs etc.) and the causal mechanisms that depend or act on those structures. The expert biologist can reason with these concepts to understand published results of other experts, generate predictions, and interpret empirical research findings. Biologists use their cognitive representations of models to generate external representations, such as verbal arguments, diagrams, graphs, and computer simulations. Biologists also continually revise models to accommodate findings that are inconsistent with predictions or to extend them to cover previously unexplained ranges of phenomena. This *model-based reasoning* is particularly central to research in biology, which involves work with intricate, interacting chains of causality.

In this light we can ask what it means to learn biology or to become a biologist. Clearly, acquiring mental models that increasingly approximate those of professional biologists and that are the basis of active model-based reasoning forms a significant part of this developmental transition. We believe that a primary goal of the large lecture course in biology should be to bring students to the point where they can reason with models in scientific inquiry, to generate predictions and interpret findings, and more generally to the point where they have acquired general skills with model-based reasoning and an appreciation of the centrality of model-based explanation in biology.

MBR-oriented instruction in large classes

The proposed project concerns the assessment practices in the "lecture" component of large university introductory biology courses. Our course, and most other such courses, also includes a laboratory component. Although our larger research and development effort includes work to integrate MBR-oriented class meetings with labs and to promote MBR in extended, hands-on laboratory exercises, this proposal concerns only the non-laboratory components of the course.

The central claim underlying MBR instruction is that learning to reason with biological models is a large part of what it means to *learn to think like a biologist*. The key goal of the instruction, therefore, is to create an environment in which students practice reasoning with and therefore learn to reason with the central theoretical models of contemporary biology. The major elements of MBR instruction are (1) MBR-specific learning cycles organized around the class meeting and the topical unit, and (2) the integration of formative and summative assessment.

The classroom learning cycle

Class meetings are devoted to the active practice of MBR with feedback and not to the review of basic material. Over the past five years we have adapted the use of in-class concept tests (Mazur, 1996; Crouch & Mazur, 2001; Landis et al., 2001) and personal response systems (PRS) (Dufresne et al., 1996; Wenk et al., 1997) to model-based reasoning in biology. Both the classroom and the topical (below) learning cycles reflect recent developments in the reform of introductory biology (Lawson, 2001; Ebert-May, 1997). There are several elements to the classroom learning cycle:

- Prior to each class students visit the course web site to answer a pre-class quiz that checks their basic understanding.
- During the class meeting the instructor poses a graded series of 2-5 questions that require model-based reasoning. The questions, which are in multiple-choice format, are projected on a large screen at the front of the lecture hall. Groups of three or four students discuss the questions. Students commit to and submit their personal answers via hand-held wireless transmitters to a central computer, which displays a histogram of the students' choices on the screen. Students receive immediate formative feedback on their progress in model-based reasoning.
- Following small-group discussion, the instructor leads a whole-class discussion in which groups explain their reasons for choosing various answers. The instructor wraps up the discussion by reflectively elaborating and repairing lines of student reasoning and summarizing the critical properties of the model that have been illustrated.
- After class the instructor posts on the website the day's MBR problems with remarks that remind students of the lessons learned about models and about scientific inquiry.

Each class meeting is a sequence of formative assessments that provide the instructor with critical information about student understanding and that serve as learning experiences for students. The problems provide structured practice (Ericsson et al., 1993) with scientific reasoning in a zone of proximal development in the specific Vygotskyan sense of extending a cognitive skill under the guidance of a more-expert practitioner (Brown, Metz, & Campione, 1996). The consistent small-group discussion establishes a community of practice in which students become comfortable with going beyond the recall of memorized facts, with trying out and improving ideas, and with verbalizing arguments in a context of scientific discourse.

The topical learning cycle

This daily formative-assessment-driven instruction is organized around larger motivating questions that integrate sequences of class meetings. An example from early in the course is, *How is it possible for organisms to have many types of cells, given that they develop from a single cell*? This question provides a framework for developing a series of models in a coherent and interconnected manner, and it sparks the organization of sequences of reasoning problems around current issues in biology, such as the possible use of stem cells in the treatment of disease.

The topical learning cycle has several elements in addition to those of the classroom cycle:

- Web-based class preparation materials and in-class instructor comments frame MBR formative learning episodes in terms of the major questions being addressed.
- Sequences of MBR problems are often couched as extended inquiries into research questions, such as therapeutic cloning. These problem sequences move well beyond textbook presentation.
- An MBR-oriented quiz is administered each week, providing a comprehensive formative assessment of the week's models.

Organizing MBR into larger topical units sustains students' motivation and insures that they understand the significance of the particular models that are introduced. Weekly quizzes challenge students to apply the models developed for a particular topic in an integrated manner, consolidating their understanding both of the individual models and of how they fit together into a larger conceptual structure. Continuous feedback from the stream of formative learning episodes during a unit guides instructors in planning class activities and informs students of their progress in learning skills of biological inquiry.

Integration and alignment of formative and summative assessment
The pre-class web quizzes, in-class discussion problems, and weekly in-class quizzes constitute a continuous stream of formative assessment that directly prepares students for the mid-term and final examinations in the course, which are heavily weighted toward reasoning with models. Students learn that they will not be able perform well on the examinations if they rely only on their memory for facts in the textbook, because the exams demand the same kind of reasoning that has been practiced in class. The in-class formative assessments also provide the course instructors with invaluable feedback. Student answers provide direct information about the general quality of reasoning in the class and trouble spots for understanding. In the short term these data allow instructors to fine tune the plans for each class, to adjust web content following each class, and to readjust the problem sequences and quizzes for a unit. The present proposal evaluates conceptual change in the classroom and seeks to develop validated summative assessments of MBR.

The portability of MBR instruction
We believe that MBR instruction in biology can be adopted at other universities. PRS technology is being widely adopted by universities. Three publishers, Prentice-Hall (a partner in this proposal), McGraw-Hill, and Benjamin-Cummings bundle PRS transmitters with texts. Web-based class support is already nearly universal in university science courses. Finally, our approach does not involve any changes in the topics or "coverage" of traditional introductory biology courses. Standard introductory textbooks can be used in MBR courses (we are currently using the 6th edition of Campbell's *Biology*). What changes is what the students learn about the topics and how the learning is assessed, which is the subject of this proposal.

The Use of Multiple-Choice Questions to Foster and Assess Model-Based Reasoning

Summative assessment in large introductory biology courses necessarily relies heavily on examinations made up of multiple-choice (M-C) questions. Students will not respond to MBR-oriented classroom instruction (or any form of inquiry-oriented instruction) unless it is effectively assessed on these M-C examinations. The proposed project therefore involves further development of and research on MBR assessment using M-C questions. Before outlining the project activities in the next section, we show here some concrete examples of how we employ M-C questions in the context of MBR instruction.

For purposes of contrast, we begin with examples of the kind of questions that we do *not* use in our course. We do not use questions such as those in Figure 1, which make up the bulk of mid-term and final exams in many introductory biology courses. Students can answer such questions (these were taken from a published test bank) via direct recall from their notes or textbooks.

Because biological models have rich combinatorial properties and apply to an unlimited range of specific situations, it is relatively easy to write questions that are not directly answered in the textbook or in class, that require model-based reasoning in novel contexts, and that involve various degrees of transfer from material that *is* presented in class, in the text, or on the course web site. Answering such questions necessarily involves factual recall, but in a context of episodes of reasoning that exercise skills of scientific inquiry.

Figure 1: Examples of recall-based multiple-choice questions that do not evoke MBR

Gene Expression

What is the proper order of events in gene expression?
a. RNA processing, transcription, translation
b. Transcription, RNA processing, translation
c. Translation, RNA processing, transcription
d. Transcription, translation, RNA processing
e. Translation, transcription, RNA processing

Which of the following occur during the initiation of translation?
a. Release factor binds
b. RNA polymerase binds to a template
c. Ribosome subunits assemble
d. Introns are removed
e. Peptide bonds form

Metabolism

During the metabolism of a single glucose molecule, which of the following produce the most ATP molecules?
a. Glycolysis
b. Pyruvate dehydrogenase
c. Krebs cycle
d. Oxidative phosphorylation
e. Fermentation

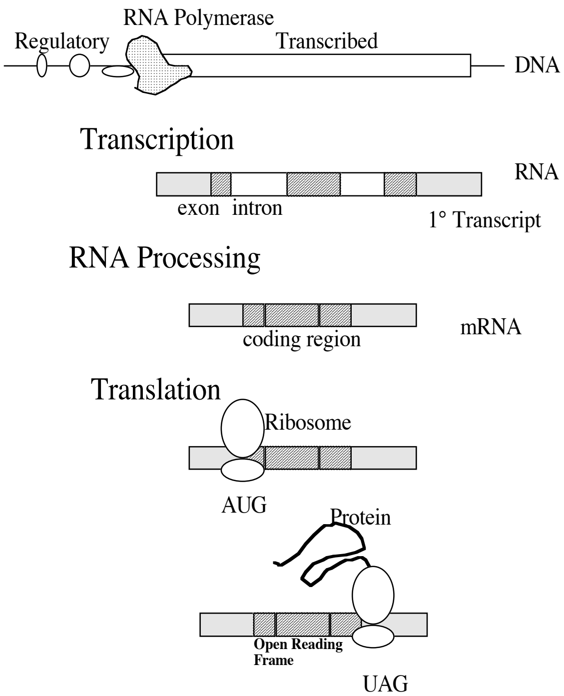
Which of the following are required by the Krebs cycle?
a. NAD+
b. GDP
c. Acetyl CoA
d. FAD
e. All of the above

Since MBR-oriented questions presented in isolation can appear to be arbitrary and unreasonably complex, it is important to remember that the questions are introduced in a meaningful context that is built up gradually and emphasizes the role of models in the cycle of scientific inquiry. Figure 2 presents a snapshot of a stage in the development of the model of gene expression during the course. This model is developed over a period of several class meetings within the topical framework provided by the question, *How can a single cell develop into a complex multicellular organism, whose many types of cells all contain identical DNA*, and motivated by the issue of the potential of stem cell research. During this topical unit students are introduced to and begin to reason about the component processes involved in gene expression, reaching a point where they should be familiar with the full *causal logic* sketched in the figure. This causal logic is a direct consequence of the set of *focal concepts* listed in the figure, and it requires the understanding of the *associated vocabulary and concepts* also listed. Students have also had substantial experience interpreting and using diagrams like that shown in the figure to augment their reasoning. The textbook and web resources serve as important resources to students in learning these concepts. An advantage of MBR questions is that reasoning with the model exercises the focal concepts and associated vocabulary repeatedly in ever-shifting contexts of application.

The explanatory role of a model in scientific inquiry defines a rich and unbounded space of reasoning problems. Questions involving a model can be graded nearly continuously in difficulty and can involve any phase of the scientific inquiry cycle. For example, problems can involve generating empirical predictions from the model, given a set of initial conditions, or given that a certain modification to the model has been made. Problems can involve using the model to explain a set of experimental results, including making modifications to the model to handle results that are inconsistent with the model's predictions. Problems can involve designing an experiment that would produce results that can discriminate between rival models. Problems can involve combining models to generate predictions concerning the interaction of their component processes, and so on. Thus, the fertility of model-based problem spaces allows sequences of MBR questions to be posed that exercise much of the reasoning involved in a full inquiry cycle, from the generation of predictions, to the design of experiments, to the interpretation of results, to the revision of the model in light of inconsistent results.

Figure 2: Gene Expression Model (See text for explanation)

Typical diagram



Causal logic (predictive & interpretive value)

Transcription produces RNA and occurs if the following conditions are satisfied:

- RNA polymerase binds
- TATA binding protein complex forms
- Stimulatory transcription factors bind to enhancers
- DNA bending occurs
- Inhibitory transcription factors don't bind, or don't contact RNA polymerase

RNA processing reduces the size of the RNA, produces mRNA, and occurs if:

- primary transcript is produced
- introns are present
- proper RNA processing molecules and RNAs are present

Translation produces protein and occurs if:

- Ribosome binds to translation start, has a start codon and ribosome assembly sequences
- Open reading frame is present

Protein size depends on length of open reading frame

Focal Concepts

- Gene structure: transcribed region, regulatory region
- Transcription factor proteins bind to DNA and regulate gene expression
  - Transcription factors regulate activity of RNA polymerase
- Cell types have particular sets of genes expressed
  - Gene regulation involves coordinated expression of particular sets of genes in each cell type
- Sets of genes respond to common transcription factors if they have similar regulatory regions
- DNA bending is required for distant transcription factors to contact and regulate RNA polymerase
- RNA processing removes sequences in DNA but not used in mRNA
- Translation uses sequences within the mRNA
- Translation occurs in open reading frame with AUG start and UAG stop
- Size of RNA and proteins can be measured with gel electrophoresis

Associated Concepts and Vocabulary

Transcription, transcription factor, enhancer / silencer, regulatory sequences, transcribed sequences, DNA, RNA, proteins, intron, exon, RNA processing, 5'/ 3' UTR, TATA box, RNA polymerase, primary transcript, mRNA, ribosome, codon, start/stop codon, reading frame, translation, cell type/identity, gene expression, gel electrophoresis, silent mutation, frameshift mutation, missense mutation, N/amino terminus, C/carboxyl terminus

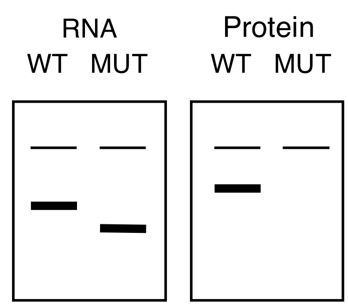
(Figure 2 continued on following page)

<u>Figure 2 continued</u>

Example Questions That Engage Model-Based Reasoning

[Question 1: Student must reason with altered versions of the model to predict a qualitative result.]
1. Which of the following is most likely to cause a slight change in protein structure?
   a. Deletion of enhancer sequences from the regulatory region.
   b. A change of a single base in the start codon.
   c. A change in RNA process that leaves an intron in the coding sequence.
   d. A change of a single base in the open reading frame.
   e. A change of a single base in the 3' UTR.

[Question 2: Student must understand data from a standard experimental method, gel electrophoresis, then must reason with alternative versions of the model to explain the data. This kind of question meshes classroom and exam experience with hands-on laboratory experience.]
2. Which of the following could explain these results obtained from a patient with a mutation in their dystrophin gene?



   a. The mutant gene has an intron that is not removed from the primary transcript. That intron has a stop codon in it.
   b. The mutant gene has a defect in its regulatory region, and silencers are absent.
   c. The mutant gene has a deletion of sequence from the 5'UTR. This change prevents the small and large ribosome subunits from binding to the mRNA.
   d. The mutant gene has a sequence change in its coding region that changes a UAC codon into a UAG codon.
   e. None of the above could produce the results shown.

[Question 3: Student must reason with the model to flesh out a hypothesis to the point where an adequate experimental design can be selected. Such questions mesh with hands-on laboratories that involve experimental design.]
3. You are analyzing a mutant gene and want to determine whether there is a defect that prevents ribosome binding. Which of the following experiments would produce data that could reject this hypothesis?
   a. Determination of the DNA sequence of the of the 5' regulatory region of the gene.
   b. Determination of the size of the mRNA produced by the gene.
   c. Determination of whether or not protein can be produced by the gene.
   d. Analysis of the mRNA for unremoved intron sequences.
   e. None of he above

The questions in Figure 2 illustrate how MBR integrates practice with the causal logic of a model and practice with scientific inquiry. The first question simply states an observation qualitatively (*slight change in protein structure*) and asks the student to reason with variants of the model to explain the observation. The second question presents schematic experimental results from a standard laboratory method (gel electrophoresis) and asks the student to explain the results. Students must understand the results (*The normal gene makes RNA and protein of known sizes, and the mutant gene makes smaller RNA and no protein*) and coordinate this understanding with their understanding of the model to reason to a modified version of the model that accounts for the results. The final question states an initial hypothesis (*defect that prevents ribosome binding*) and asks the student to flesh the hypothesis out in terms of the model to the point where a sound experimental design to test the hypothesis can be selected.

The Theoretical Case for MBR as a Primary Learning Goal and the Challenges of Implementation

Conventional lecture-based courses have significant shortcomings. Most students in introductory college biology learn facts about biological models but do not learn to reason with the models. They therefore practice the skill of memorizing relatively isolated facts more than they practice the processes of scientific reasoning. They tend to think of biology as a collection of things we know rather than as an active body of theory that evolved and continues to evolve through the quest to explain phenomena. They rapidly forget the facts they have memorized, because the facts are not integrated into meaningful conceptual structures (See Anderson, 2000, chapters 6-7 for the influence of internal mental organization on memory).

Organizing the learning environment around reasoning with biological models (MBR) has potential advantages. Learning to reason with the key explanatory models in a science promotes, indeed largely constitutes, the understanding of the science (For relevant research on physics see e.g. McDermott & Redish, 1999; Clement, 1993; White & Frederiksen, 1990; Bao & Redish, 2001; for chemistry see e.g. Harrison & Treagust, 1996; Williamson & Abraham, 1995; for biology see e.g. Modell, 2000; Michael et al., 2002; Stewart et al., 1992; Cartier, 2000; Lawson, et al., 2000; for technology-enhanced learning environments see e.g. Raghavan et al., 1998; Gobert, 2000; Gobert et al., 2002). Organizing knowledge in terms of explanatory models and reasoning is a way of organizing knowledge in terms of contexts of use, which is known to promote long-term retention and retrievability (Anderson, 2000, chapters 6-7; Bransford, Brown, & Cocking, 2000, chapters 2-3). Focusing on reasoning with concepts, particularly in multiple problem contexts, facilitates the transfer of learning to novel situations (Bransford & Stein, 1993; Bransford, Zech, et al., 2000). Instruction in and practice with reasoning facilitates the development of metacognitive strategies, which also facilitate transfer (Bielaczyc et al., 1995; Kuhn, 1991; Scardamalia et al., 1984). Finally, and perhaps most importantly, students who are actively reasoning with biological models are on a developmental trajectory toward the acquisition of skills that are an important component of expertise in the biological sciences.

Organizing assessment around MBR in large courses. Our revision of introductory biology has brought us to the threshold of three significant assessment issues: (1) The set of MBR items that we have developed must be expanded and content-validated; (2) Student gains in model-based reasoning during the course must be confirmed; (3) Methods for developing reliable, valid, and practical summative assessments of individual student performance on MBR must be developed. These issues are addressed in the research and development plan described below.

Refinement and Validation of Assessment Practices for Reasoning with Biological Models

Over the past three years we have developed MBR-oriented instruction to the point where the course can serve as an effective test bed for the work proposed here: the further development, validation, and dissemination of formative and summative assessment practices that are inherent to the instruction. In this section we outline the research and development plan for the project. We first describe the *assessment data* that will be collected during the project, as well as the *methods* underlying each data source. Then we describe four *assessment research projects* that form the core of the proposal: (1) Validation and refinement of problem characteristics and quality; (2) Group-level studies of student progress; (3) Quasi-experimental comparison of MBR vs. non-MBR oriented instruction; and (4) Construction of individual summative assessments for MBR.

Assessment Data and Methods

The proposed assessment research employs a range of methods, each of which delivers a stream of data that is employed in one or more of the research projects described in succeeding sections:

- Expert judgments of problem content and difficulty. The project staff categorizes MBR problems according to their content and rates their degree and type of reasoning challenge. In the proposed project additional problem ratings will be collected from an outside group of biology teachers.
- Student answers to multiple-choice items. Students answer hundreds of multiple-choice (M-C) questions per term in both formative and summative assessment contexts, and all answers are electronically recorded and indexed by student identification number.
- Analysis of student essays. We will score samples of student essays taken from pure essay questions or hybrid multiple-choice-essay questions (in which students are asked to write explanations of their M-C selections). A previously-developed reasoning components rubric will serve as the basis for this work (Izumi, 2003; Rea-Ramirez & Izumi, 2003).
- Videotapes of student discussion. Whole-class and small group discussions are periodically videotaped, allowing analysis of student reasoning in class.
- Data on student characteristics. Information on student demographics, SAT scores, university grade-point averages, and so on will be available during the project, with appropriate IRB oversight. We also periodically collect information on student attitudes toward MBR-oriented instruction, which can be entered into data analyses.

Project 1: Validation and refinement of problem characteristics and quality

This strand of project research is designed to insure that the problems employed in the course are well aligned with expert judgments concerning model-based reasoning, succeed in evoking reasoning in students, span a wide range of difficulty, and are free of basic psychometric flaws.

Expert ratings of reasoning demands
In addition to judgments of problem type and difficulty made by the project staff, a group of college biology instructors from outside the project will rate a large sample of the assessment items used in the course, ranging from those requiring straight factual recall to those requiring challenging model alterations or combinations. First, the experts will be asked to rate the similarity of pairs of test items according to the knowledge and skills measured by the items. These item similarity ratings will be analyzed using multidimensional scaling to determine the content and cognitive characteristics that the experts used in making their similarity judgments (Sireci, 1998; Sireci & Geisinger, 1995). Our basic hypothesis is that MBR items will be distinguished from the factual recall items in the multidimensional perceptual space derived from the experts' ratings. Second, the experts will be asked to rate all items according to their cognitive complexity. MBR items are predicted to receive higher ratings on the complexity scale than the recall-oriented items. In both rating tasks, the experts will be blind to the intended constructs measured by the items. The experts will also be asked to comment on the items, allowing us to identify items that are poorly written, and to further refine our own analytical judgments of the reasoning challenges of the problems.

Feedback from formative assessments
The results of formative assessments, including in-class MBR problems, essays, and web quizzes, provide detailed information about student reasoning that will be used to improve the reasoning problems and problem sequences that drive the instruction. From in-class videotapes and student essays instructors and researchers can see fairly directly whether a problem evoked the lines of reasoning in students that were anticipated. Misconceptions or gaps in the understanding of a model are relatively easy to identify, and problems can be revised, or new problems written, to move students to work through the

misunderstandings. Confusing problem statements (question stems) or poorly-written M-C alternatives are also relatively easy to identify and repair. The use of formative assessment outcomes to revise problems, problem sequences, and surrounding instructional materials will occur throughout the project.

Multiple-choice item selection and calibration
Psychometric approaches will be used to select, refine, and calibrate problems that have been presented in straight M-C format on either formative or summative assessments (quizzes, mid-term exams, or final exams). Classical item statistics are available routinely for M-C exams and can be used to estimate the difficulty level of problems and to help identify defective items. We believe that our sample sizes are large enough to investigate the use of item response theory (IRT) to calibrate item difficulties for the construction of assessments as well. IRT will allow us to place all item formats onto a common scale, even when different samples of students took different sets of items. Calibrating items using IRT will also enable us to measure change in students' performance over time, without re-administering the same items to students.

Outcomes of the problem quality research

- Description of practical techniques for insuring problem quality in MBR-oriented teaching.
- Problems and problem sequences that are well written, have known difficulty, and that lead students to work through known misconceptions and sources of difficulty.
- Problems with a wide range of difficulty that can be used in summative assessments.

Project 2: Studies of student progress

During the first two years of the project we will undertake research to show in detail that student reasoning develops over the course of MBR-oriented instruction. This work will extend previous research in which we analyzed student essays written at the beginning and end of the semester and coded videotapes of peer and whole-class discussions of model-based problems (Rea-Ramirez & Izumi, 2003; Stillings, 2002).

Use of essays to study changes in students' model-based reasoning during the course
The goal of this strand of the project is to employ essay or M-C-essay hybrid assessments, scored according to a reliable rubric, to demonstrate student progress on the reasoning components. This research is intended to show that MBR-oriented, assessment-driven instruction leads to increases in student reasoning performance. The prediction is that increasing numbers of students will display critical components of reasoning (e.g. increasing use of models, increased selection of appropriate models, improved model-to-situation mapping, improved model application, and improved model revision) and that performance within these components will also improve (e.g. higher scores within model application). The proposed research makes use of a variety of pre-post comparisons:

Change within a topical unit. For selected topical units students will be given essay assessments near the beginning and at the end of the unit. Our prediction is that there will be increases in the quality of reasoning during a unit.

Change from unit to unit. We will compare essay performance at the beginning of a unit that comes early in the course with performance at the beginning of a unit that comes later in the course. Our prediction is that as the course progresses students will begin their study of a new set of models with stronger reasoning skills.

Change in transfer to novel models. At the beginnings and ends of selected terms we will give essay problems involving models that are not covered in the course. The basic prediction is that there will be a significant pre-post increase in reasoning scores for the class as a whole.

In addition to examining change for entire classes, the relationship between student characteristics (e.g. grade-point average, sex, or previous exposure to biology) and the above types of change will be examined. We have established in previous work that student characteristics are not related to attitudes toward the MBR reform or to final grades before vs. after the reform. These results suggest that MBR has not had a global differential impact on different populations of students, but they say nothing about the relationship between progress on reasoning skills and student characteristics, which will be directly examined in this project. From the standpoint of instructional design, our goal is to structure the formative assessments and other aspects of the course in such a way that students with a wide range of reasoning skills at the beginning of the course will be able to make progress.

Several methods will be available to equate pre- and post-tests: (1) If sufficient data are available, IRT methods can be used calibrate all items onto a common scale using data from previous terms; (2) If necessary, some problems will simply be repeated on the pre- and post-tests; (3) If necessary, the class will be split randomly into two groups which will receive the same pair of problems in different order on the pre- and post-tests (e.g. Group 1 receives Problem A on the pre and Problem B on the post, while Group 2 receives Problem B on the pre and Problem A on the post).

Essays will be scored using the procedures and the reasoning rubric developed in earlier research. Scorers train on a common sample of essays from previous research, and a substantial sample of the research essays are scored by two researchers to check for inter-scorer reliability. Scorers are blind to the pre-post status of the essays, which is also not identifiable from question content.

Use of videotapes to study changes in students' model-based reasoning during the course
In recent work we have used our methods for scoring student reasoning to analyze changes in students' reasoning during whole-class and small-group discussion during the semester. In an analysis of a sample of videotaped whole-class discussions during a term we found that student use of models improved during the term and that instructor scaffolding was steadily reduced (Rea-Ramirez & Izumi, submitted). This analysis was an initial confirmation that reasoning during student discussion actually improves during the course. In the proposed research we will replicate and extend this analysis.

Outcomes of studies of student progress

- Test of the hypothesis that student reasoning skill improves during instruction on a topic.
- Test of the hypothesis that advances in reasoning skill transfer across topics.
- Test of the hypothesis that advances in reasoning skill transfer to unfamiliar models.
- Test of the hypothesis that student reasoning in small-group and whole-class discussion improves during the course.
- Test of the hypothesis that the expert rating data predict MBR reasoning content and difficulty.
- Further refinement of problem content and problem sequencing based on assessment evidence.

Project 3: Quasi-experimental comparison of MBR vs. non-MBR oriented instruction

During the project we will run a traditionally-taught section of introductory biology as a comparison group to the MBR section. The comparison section will be taught as a traditional lecture class, covering the same material and including the same laboratories as the MBR class. The two sections will have different instructors with comparable experience and student ratings. The two sections will be statistically comparable on relevant student characteristics.

Essay-based comparison of changes in reasoning skill
The essay-based assessments described above will be administered in the comparison section, although in a given semester the comparison section will receive only one pre-post pair of essay assessments. The assessments will be incorporated into the course and presented by the instructor as regular course work

with students receiving the same amount of credit assigned in the MBR section. Essay scorers will be blind to the section from which essays came.

The basic hypothesis underlying this study is that students in the MBR section will show greater improvements in reasoning skill than students in the comparison section. In different semesters this hypothesis will be explored with respect to gains within a topic, gains from topic to topic, and transfer to unfamiliar models, as outlined above. In addition the differences between the MBR and comparison sections will be explored with respect to student characteristics. Our scoring rubrics will allow the reasoning differences between the MBR and comparison groups to be explored in considerable detail.

Comparisons based on multiple-choice summative assessments
Mid-term and final examinations for the MBR and comparison sections will have overlapping content on questions ranging from those requiring simple knowledge retrieval to those requiring considerable model-based reasoning. Our basic hypothesis is that MBR students will show superior performance on reasoning-intensive items. The sensitivity of differences between the groups to item content and to student characteristics will also be assessed.

Outcomes of MBR vs. non-MBR comparisons
- Well-controlled test of the hypothesis that MBR instruction improves students' model-based reasoning skills more than traditional lecture-based instruction.
- Detailed investigation of the differential effects of MBR instruction on components of reasoning.
- Well-controlled test of the hypothesis that the effects of MBR instruction can measured with multiple-choice summative assessments.

Project 4: Construction of summative assessments for MBR

In the studies described above, we propose to use essay-style assessments to demonstrate that the quality of model-based reasoning over an entire class of several hundred students improves during assessment-driven, MBR-oriented instruction. Essay-style assessments will be used in this research because they allow relatively direct assessment of model-based reasoning, but they are impractical to employ routinely for the entire content of a mid-term or final examination in most large introductory biology courses. Thus, a further goal of the project is to demonstrate that valid, largely multiple-choice-based summative assessments of individual students' model-based reasoning abilities can be constructed under typical university instructional conditions. Out of concern for practicality, we emphasize straightforward methods of achieving this goal.

<u>Systematic use of expert ratings of problems</u>. The judgments we collect from a panel of expert teachers will be correlated with student performance on the same problems when they are subsequently employed in formative or summative assessments. We hypothesize that biology teachers' ratings of the reasoning demands of problems will be reasonably accurate. If this hypothesis is borne out, local instructional staffs will be able to learn to use their own judgments systematically to construct examinations with varied reasoning challenges.

<u>Formative item trials</u>. The integration of formative and summative assessment in MBR-oriented instruction allows multiple-choice items to be tested in formative contexts prior to use on summative examinations in a later version of the course. Formative pre-trials involve relatively low stakes for students and can deliver rich information about student reasoning without having to assign a score to every student. For example, the quality of the student discussion in response to a concept test can be immediately monitored by the instructor and teaching assistants, and the quality of student reasoning on an M-C-essay hybrid question can be determined by reading a sample of the responses. Over the course

of the project items that demonstrated varying degrees of reasoning challenge in formative contexts will be banked for use on summative examinations in later terms.

Inclusion of essay items on summative assessments. Some M-C-essay hybrid items will be included on summative examinations. The essays allow a direct check on whether students who select the correct alternative can produce lines of supporting reasoning. Correlations of the essay scores with subscores on collections of items with varying levels of difficulty will provide additional evidence concerning the validity of largely M-C based examinations.

Use of item statistics. Item statistics will be used to confirm that items have behaved as predicted by expert ratings and formative pre-tests. Within a topic area, the empirical item difficulties should accord with the predicted difficulties. Failures will suggest changes in one or more items.

Dimensionality of assessment data. Several approaches will be employed to investigate the dimensionality of student performance on summative assessments, including exploratory and confirmatory factor analysis, multidimensional scaling analysis, and item response theory (IRT) analysis. Hypothesized structures will be derived from expert item ratings and analyses of student reasoning on essays. If a stable dimensional structure emerges from this research, it will be used to refine summative assessments over the course of the project and to produce a set of guidelines for writing problems that are likely to load clearly on particular dimensions.

Implications of the research on summative assessment

The research on summative assessment addresses the pervasive problem of aligning formative and summative assessment in very large reform-oriented science courses. We hope to demonstrate that reliable, valid, and practical summative assessments can be constructed and refined through the use of expert ratings, formative item trials, limited student essays, and item statistics. These techniques are available to any department running a large course, and they allow courses to vary in their specific content and reasoning demands. The synergy between formative and summative assessment is the key element in our approach. Through the continuous use of MBR-oriented formative assessment, students learn what it means to reason with models, and faculty members are able to evaluate specific items by observing detailed student reasoning behavior and to develop their intuitions about the difficulty levels of different types of items. Extensive formative assessment allows faculty members to construct reasonable multiple-choice-oriented summative assessments, and it allows students to approach those assessments with the right cognitive orientation.

References

Advisory Committee to the National Science Foundation (1996). *Shaping the Future: New Expectations for Undergraduate Education in Science, Mathematics, Engineering, and Technology*. NSF 96-139. Washington, D.C.: National Science Foundation.

American Association for the Advancement of Science: Project 2061 (1993). *Benchmarks for Science Literacy*. New York: Oxford.

Anderson, J. R. (2000). *Cognitive Psychology and Its Implications, 5th ed*. New York: Worth Publishers.

Bao, L. & Redish, E. F. (2001). Model Analysis: Assessing the Dynamics of Student Learning. Preprint, University of Maryland.

Bielaczyc, K., Pirolli, P., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13, 221-252*.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (2000). *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. Washington, DC: National Academy Press.

Bransford, J. D., & Stein, B. S. (1993). *The IDEAL Problem Solver, 2nd Ed.* New York: Freeman.

Bransford, J. D., Zech, L., et al. (2000). Designs for environments that invite and sustain mathematical thinking. In P. Cobb, E. Yackel, & K. McClain (Eds.) *Symbolizing and Communicating in Mathematics Classrooms: Perspectives on Discourse, Tools, and Instructional Design*. Mahwah, NJ: Erlbaum

Brown, A.L., Metz, K., and Campione, J.C. (1996). Social interaction and individual understanding in a community of learners: The influence of Piaget and Vygotsky. In Tryphon and Voneche (eds.), Piaget-Vygotsky: *The Social Genesis of Thought,* 145-170. Pyschology Press.

Cartier, J., Assessment of explanatory models in genetics: Insights into students' conceptions of scientific models. (2000) Research Report - National Center for Improving Student Learning and Achievement in Mathematics and Science. Report No. 98-1

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5, 121-152*.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching, 30(10), 1241-1257.*

Crouch, C. H. and Mazur, E. (2001). Peer Instruction: Ten years of experience and results. American Journal of Physics, 69, 970-977.

Dufresne, R. J., Gerace, W. J., Leonard, W. J., Mestre, J. P., & Wenk, L. (1996). Classtalk: A classroom communication system for active learning. *Journal of Computing in Higher Education, 7(2), 3-47*.

Ebert-May, D. & Brewer, C. (1997). Innovation In Large Lectures-Teaching For Active Learning. *Bioscience*, *47(9)*, 601-608.

Ericsson, K. A., Krampe, R. T., Tesch-Roemer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406.

Gentner, D., & Stevens, A. L. (Eds.) (1983). *Mental Models*. Lawrence Erlbaum Associates.

Gobert, J. (2000). A typology of models for plate tectonics: Inferential power and barriers to understanding. *International Journal of Science Education, 22(9),* 937-977.

Gobert, J., Snyder, J., & Houghton, C. (2002). The influence of students' understanding of models on model-based reasoning. Presented at the Annual Meeting of the American Educational Research Association, New Orleans, LO, April 1-5.

Harrison, A. G. & Treagust, D. F. (1996). Secondary students’ mental models of atoms an molecules: Implications for teaching chemistry. *Science Education, 80(5), 509-534*.

Hempel, C. G. (1965). *Aspects of Scientific Explanation*. New York: Free Press.

Izumi, A. (2003). Measuring measurements of scientific thinking: Do multiple forms of assessment exhibit similar evidence of student understanding on complex scientific reasoning problems. *Doctoral dissertation*, School of Education, University of Massachusetts, Amherst.

Kuhn, D. (1991). *The Skills of Argument*. Cambridge University Press.

Landis, C. R., Ellis, A. B., Lisensky, G. C., Lorenz, J. K., Meeker, K., & Wamser, C. C. (2001). *Chemistry ConceptTests: A Pathway to Interactive Classrooms*. Prentice-Hall.

Lawson, A. E. (2001). Using the learning cycle to teach biology concepts and reasoning patterns. Journal of Biological Education, 35(4), 165-69.

Lawson, A. E., Clark, B., Cramer-Meldrum, E., Falconer, K. A., Sequist, J. M., Kwon, Y-J.. (2000). Development of Scientific Reasoning in College Biology: Do Two Levels of General Hypothesis-Testing Skills Exist? *Journal Of Research In Science Teaching,* 37(1), 81–101.

Magnani, L., Nersessian, N. J., & Thagard, P. (Eds.). (1999). *Model-Based Reasoning in Scientific Discovery*. Kluwer Academic Publishers.

Mazur, E. (1996). *Peer Instruction: A User's Manual*. Prentice Hall.

McDermott, L. C., & Redish, E. F. (1999). Resource letter on Physics Education Research, *Am. J. Phys. 67 (9) 755*.

Michael, J. A., Wenderoth, M. P., Modell, H. I., Cliff, W., Horwitz, B., McHale, P., Richardson, D., Silverthorn, D., Williams, S. & Whitescarver, S. (2002). Undergraduates' understanding of cardiovascular phenomena. *Adv. Physiol. Educ. 26:72-84*.

Modell, H. I. (2000). How to help students understand physiology? Emphasize general models. *Adv. Physiol. Educ. 23:101-107*.

National Research Council (1996). National Science Education Standards. Washington, D.C.: National Academy Press.

Nersessian, N. J. (2002). The cognitive basis of model-based reasoning in science. In Carruthers, P. & Stich, S. (Eds.), *The Cognitive Basis of Science*, pp. 133-153. New York: Cambridge University Press.

Pellegrino, J., Chudowsky, N., and Glaser, R. (2001). *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: National Academy Press.

Raghavan, K., Sartoris, M. L., & Glaser, R. (1998). Why does it go up? The impact of the MARS curriculum as revealed through changes in student explanations of a helium balloon. *Journal of Research in Science Teaching, 35* (5), 547-567.

Rea-Ramirez, M. & A. Izumi (2003). Empirical Evidence On Student Understanding Of Biological Models: Fat Metabolism. *Presented at the Annual Meeting of the National Association for Research in Science Teaching*, Philadelphia, March 23-26, 2003.

Rea-Ramirez, M. & A. Izumi (submitted). Instructor Scaffolded Model Construction and Model Based Reasoning in a College Introductory Biology Class.

Scardamalia, M. C., Bereiter, C., & Steinbach, R. (1984). Teachability of reflective processes in written composition. *Cognitive Science, 8, 173-190.*

Singley, M. K., & Anderson, J. R. (1989). *The Transfer of Cognitive Skill*. Harvard University Press.

Sireci, S.G. (1998). Gathering and analyzing content validity data. Educational Assessment, 5, 299-321.

Sireci, S.G., & Geisinger K.F. (1995). Using subject matter experts to assess content representation: An MDS analysis. Applied Psychological Measurement, 19, 241-255.

Stewart, J., Hafner, R., Johnson, W., and E. Finkel (1992). Science as model building: Computers and high school genetics. Educational Psychologist 27:317-336.

Stillings, N. (2002). Advancing our understanding and assessment of student learning through classroom-oriented research partnerships. Presented at *Pathways to Change: An International Conference on Transforming Math & Science Education in the K16 Continuum*, Washington, DC, April 18-22, 2002.

Thagard, P. (1988). *Computational Philosophy of Science*. MIT Press.

Wenk, L., Dufresne, R., Gerace, W., Leonard, W., & Mestre, J. (1997). Technology-assisted active learning in large lectures. In A.P. McNeal & C. D'Avanzo (Eds.), *Student-active science: Models of innovation in college science teaching* (pp. 431-451). Philadelphia, PA: Saunders College Publishing.

White, B. (1999). The Red and White Yeast Lab: an introduction to science as a process. American Biology Teacher 61(8): 600-604.

White, B. & Frederiksen, J. (1990). Causal model progressions as a foundation for intelligent learning environments. *Artificial Intelligence, 42, 99-157.*

White, B. & Frederiksen, J. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction, 16, 3-118.*

Williamson, V. M. & Abraham, M. R. (1995). The effects of computer animation on the particulate mental models of college chemistry students. *Journal of Research in Science Teaching, 32(5), 521-534.*