ADVANCING OUR UNDERSTANDING AND ASSESSMENT OF STUDENT LEARNING THROUGH CLASSROOM-ORIENTED RESEARCH PARTNERSHIPS

NEIL STILLINGS
Center for Research in Education and Learning[REAL]
& School of Cognitive Science
Hampshire College, Amherst, MA 01002
nstillings@hampshire.edu

The higher-order cognitive outcomes in science learning that are increasingly sought by faculty members form a complex network of concepts and skills. Analysis of the theoretical and empirical bases for these constructs in cognitive psychology suggests that they are attainable but very difficult to achieve, because of their abstract nature, the crucial role of metacognition, and the involvement of multiple cognitive systems. This suggests, first, that higher-order cognitive outcomes should be a goal of science instruction at all levels, in all disciplines, and in all courses, and, second, that the search for and refinement of appropriate strategies of instruction and assessment requires a continuous integration of curriculum development and research. The assessment of both learning environments and students' achievement of higher-order cognitive outcomes has an intrinsic research dimension. Teachers can not simply passively adopt current "best practices." Advances in science teaching and learning require that they become active and reflective researchers in cognition and education. Further, abstract theories of learning and instruction result in effective practice only if research is conducted in the learning environments in which the theories are being applied. A consequence is that sustained partnerships between teachers and researchers are a key element in the continuing effort to understand the dynamics of science teaching and learning and to design appropriate environments and assessments. The author's research team has developed a model for such partnerships, which are illustrated with a particular case involving introductory biology instruction at a large university.

Presented at *Pathways to Change: An International Conference on Transforming Math & Science Education in the K16 Continuum*
Washington, DC, April 18-22, 2002

This paper is based in part upon work supported by the National Science Foundation under Grants No. 9720363 and 9980519. The opinions, findings, and conclusions or recommendations expressed here are those of the author and do not necessarily reflect the views of the National Science Foundation. The author acknowledges the contributions of Laura Wenk, Mary Anne Ramirez, Loel Tronsky, Alisa Izumi, Samia Khan,, Carol Smith, Steve Goodwin, Randy Phillis, & Steve Brewer to the ideas and findings reported here.

Introduction

Increasing numbers of faculty members in all disciplines, in all types of institutions, teaching at all levels of instruction have adopted higher-order learning goals for all students. Many have also shown a willingness to organize instruction in new ways to achieve these goals. For those of us who are committed to a broader range of learning goals and teaching methods, the ground has shifted in recent years. The goal of creating a foundational group of faculty members, departments, and institutions that are active in experimentation and reform has been achieved. The challenge now is to build a new, national culture of undergraduate science instruction on this foundation. Building this culture requires the transmission of a set of values and commitments to faculty members throughout higher education, and it requires further institutionalization at the national level and within colleges and universities. These changes now hinge, in turn, on vigorously continuing to investigate, refine, evaluate, document, and disseminate the range of changes that are in varying stages of implementation across the country. In this paper I describe the approach to these issues that we at the REAL[1] Center have developed in our research on higher-order outcomes in science education.

Focus on learning

I begin with a point that is often somehow forgotten or slighted in the heat of discussions about curriculum or course design: It is important to articulate intended learning outcomes separately from methods of instruction. Both traditional and reform-minded faculty members can become wedded to particular methods, losing sight of the point that methods of instruction are not ends in themselves. They serve learning goals and should therefore be adopted, modified, and evaluated relative to those goals. In addition, the content of a curriculum is not self-evidently related to learning goals or outcomes. Depending on how it is taught, the same content can serve a range of goals, and the mere presence of particular content does not guarantee any particular learning outcome. Reflection on what it is that we want students to learn must inform instructional design, implementation, research, and evaluation at every point.

Higher-order learning outcomes

One, perhaps not so obvious, reason for clarifying learning goals is that the goals themselves must be refined or revised over time. Society's needs, scientific methods and theory, and theories of learning all change, and our learning goals must reflect those changes.

The focus of this paper is higher-order cognitive outcomes in science learning. The term *higher-order* will be used overly broadly in this paper to refer to outcomes that significantly involve the transfer of learning to substantially novel situations or to the acquisition of reasoning skills, communication skills, or systems of understanding that have a significant degree of domain independence.[2] Higher-order cognitive outcomes have been a theme in educational theory since antiquity and have played a prominent role in American discussions of education for the last 100 years. Current discussions and research have been shaped by at least three factors. First, the increasing importance of science in society suggests the need for ordinary citizens who can think independently about science. Second, the rapid pace of scientific change, as well as changes in our understanding of the nature of science, suggest that, in addition to learning current bodies of knowledge, students should be learning how to learn science, how to think scientifically, and how scientific research works. These first two factors concern the desirability of higher-order outcomes and seem to me to be unlikely to change. A third factor concerns changes in the science of learning itself. The cognitive revolution in psychology and neuroscience has changed the way we think about and study higher-order cognitive outcomes. This third factor, and its implications for the first two, is at the heart of the present paper.

Interviews with science faculty members reveal a stable taxonomy of desired higher-order cognitive outcomes. Although the relative emphasis varies, there is a high level of agreement about the list of outcomes across faculty members who teach in different settings and employ different instructional and assessment practices. These endorsed outcomes constitute part of an implicit or vernacular learning theory that is subject to empirical evaluation and theoretical revision.

COGNITIVE SKILLS

Some outcomes endorsed by faculty are cognitive skills in the sense that students are expected to acquire some of the active skills of a working scientist at a rudimentary level. Several interrelated clusters of cognitive skills have emerged in our analyses. A critical issue is the extent to which these skills are intended to transfer beyond the context and discipline of instruction.

Inquiry

Several of the skills are related to the cycle of inquiry that is typical of scientific research and that is often cited as a foundation of reform-oriented curricula: Question-theory-hypothesis formation (often within a theoretical framework); research design; data gathering; data analysis & interpretation; theory-hypothesis critique and reformulation. The curricular goal is that students will learn to generate and recognize researchable questions, to generate and critique research designs, to organize and analyze data, and to think through the implications of data for hypotheses or theories.

Quantitative thinking

Inquiry skills have quantitative dimensions that are often singled out for special comment by faculty members. Although these skills are closely related to the inquiry skills, faculty members distinguish them conceptually and often describe specific instructional strategies for improving them. One set of skills concerns handling quantitative data. Students are expected to learn how to create and interpret graphical and tabular summaries of data sets and to reason about variability and error in data. Another set of skills involves setting up and reasoning with quantitative models, either to make predictions or to interpret data.

It is important to note that faculty members distinguish these skills from skills at symbol manipulation and calculation. The latter skills may be important instructional goals, but they are not typically discussed in the context of higher-order learning outcomes. The inquiry-oriented quantitative skills might be described as a general ability to think about data and models and their relationship to reality. For example, instructors might distinguish between being able to calculate the standard deviation of a set of values and being able to think about the relationship between sample size and variability, or between being able to take the first or second derivatives of a symbolic expression and being able to go from a verbal hypothesis to an appropriate graphical sketch of the relationship between two variables.

Reading primary literature
In some curricula, teaching students to locate, organize, and read primary literature is a distinct goal (although the accessibility of the literature in particular research areas influences the degree to which this goal is adopted for beginning students). Since reading primary literature intelligently requires inquiry skills, there is considerable overlap between the two areas. Faculty members list additional skills, however, that they believe are involved in working with primary literature. These skills include locating relevant literature, ranking papers by their sources and abstracts, quickly identifying important claims, and generating a conceptual organization for a set of papers. They depend on developing an understanding of the discourse of scientific writing, which ranges from knowing how scientific papers are structured to being sensitive to how researchers speculatively extend the generality of their findings in discussion sections. More generally, exposure to primary literature may help students to understand and learn to cope with the conflicting findings and disagreement that are typical of active areas of research.

Scientific communication
A final cluster of skills concerns scientific communication, or more direct participation in scientific discourse. These skills include working collaboratively, commenting on others' work, and presenting one's own ideas in writing, in conversation, and in stand-up presentations.

STUDENTS' EPISTEMOLOGIES

Faculty members often mention a desire to change students' views about the nature of science and the status of scientific knowledge. We have made a distinction between philosophical and sociological concerns, although these two categories are not always easy, or even possible, to separate.

On the philosophical side faculty members expressed a hope that students would shift from relatively naive to more sophisticated views of science. In a generic naive view science is a body of unassailable knowledge possessed by scientific authorities, and the scientific method is a fixed procedure that can be applied piecemeal to natural phenomena to yield proven truths. More sophisticated views recognize the uncertain, open-ended, and theory-laden nature of scientific knowledge, and they show a more nuanced appreciation of the uncertainties, details, and importance of the research process and the evidence it produces.

On the sociological side the hope is that students will move toward a deeper appreciation of the social nature of the scientific enterprise. Scientific research is carried out within a social context by groups of researchers and it is therefore influenced by a complex of historical, cultural, political, and social factors. Further, the pursuit of scientific research and development has political and moral dimensions.

GENERALITY VS. FIELD SPECIFICITY OF HIGHER-ORDER OUTCOMES

The above higher-order outcomes are distinguished by their potential independence from disciplinary content. The hope is that students will improve on these dimensions not just in the immediate field of study but in general. For example, students in an appropriately-taught biology course would improve in their ability to evaluate experimental designs not only in biology but in other fields.

The learning of established, field-specific content has traditionally been the primary goal of college science instruction, and it remains a high priority. Although it is reasonable in some respects to be concerned about trade-offs between field-specific and higher-order outcomes, it is perhaps more to the point to realize that the two are not distinct instructional goals. First of all, faculty members have increasingly adopted active concept mastery as opposed to more memory-based learning. The ability to reason with the theories and causal models of a field is itself a higher-order cognitive outcome, and it may transfer to other disciplines. Secondly, disciplinary concepts potentially change their character even further when they are acquired in the context of higher-order cognitive skills.

The theoretical status and achievability of higher-order cognitive outcomes

From a common-sense perspective the list of outcomes above is complex and extremely ambitious. In some respects, the target outcomes look even more daunting from the standpoint of theories of cognition and learning.

COGNITIVE COMPLEXITY OF THE OUTCOMES

First, the outcomes involve multiple representational systems that are supported by different brain regions, at least including language, visual perception and imagery, mathematics,[3] deductive reasoning, and executive control. Most of the individual skills involve the coordination of cognitive systems. For example, interpreting a graphical presentation of data might involve reasoning that a theory predicts the slowing of a rate of change over time, imagining how a graph depicting that relationship might look, and checking if one's visual image can be adjusted to match one's perception of the graph as it is depicted on the page. The cognitive complexity of the desired outcomes alone suggests both extended practice and carefully sequenced learning environments.

ABSTRACT NATURE OF THE OUTCOMES

Second, the outcomes are abstract. Students are being asked to recognize features of situations that are independent of much of their specific content. For example, experimental designs that involve confounded variables have common abstract structures[4] that can be instantiated by an unlimited variety of content. Achieving these abstractions is certainly a difficult problem in transfer of learning, and there is some controversy about whether fully abstract reasoning structures are learnable at all [8] [9]. Two somewhat different cognitive mechanisms have been proposed that could support abstract transfer. The first is analogical mapping, in which features in a known case or domain are mapped onto the problem domain. For example, a student could think through a problem in experimental design in exercise physiology by recognizing its similarity to a previously-experienced experimental design in plant pathology. Specific applications of analogical transfer do not have to be learned in advance, but the process can not function unless the learner has appropriate knowledge outside the current problem domain, and it may also depend on learning to generate analogies, a relatively unexplored metacognitive skill. The second mechanism involves the development of abstract schemas that can be instantiated by a large range of particular content. For example, just like $y=mx+b$ is an abstract schema in which the variables can be filled with particular values, there might be a cognitive schema for experimental design that contains variables for *experimental condition* and *control condition*, which could be applied to a range of problems in experimental design. There is evidence that such schemas can be learned [11; 12], as well as evidence that they are not learned in traditional instruction [9; 10]. A great deal of research remains to be done, however, on the conditions of instruction and learning that promote the development of either analogical or schema-based transfer.

ROLE OF METACOGNITION

Third, the outcomes require *metacognition,* that is, the ability to reflect on the status of one's own concepts and theories [10] and to monitor and control complex trains of reasoning. The key, top-level outcome is the ability to coordinate ones epistemology with specific reasoning skills. In situations involving explanation a metacognitively active epistemology will lead students to ask themselves: What is the explanation at issue here; what is the theoretical framework that supports the explanation; are there alternative explanations; does the available evidence favor the explanation over others; what other evidence might be brought to bear; and so on. It is essential that an epistemology be organized in such a way that appropriate situations trigger this kind of abstract reasoning schema. However, once the schema is triggered, further reasoning is dependent on linkages to particular reasoning skills. An ability to reason about experimental designs might be required to assess the available evidence in a situation, for example. The role of metacognition in human cognitive development is well established [13]. There is also evidence that epistemological development continues well into the life span [14; 15].

Implications for Education

NEED FOR RESEARCH

The need for continuing research is underscored by this inventory of the desired higher-order outcomes for science learning and their status relative to current theories in the cognitive and learning sciences. There is evidence that the outcomes are achievable, but they are very complex and currently not widely distributed in the population of adults. There is much to be learned at a theoretical level about how higher-order outcomes are represented and acquired and about how much generality can actually be achieved. At the applied level there is much to be learned about how to foster higher-order outcomes in particular disciplines and in particular institutional settings.

NEED FOR CONSISTENT CURRICULAR INFUSION

The evidence suggests that a commitment to higher-order outcomes must be infused into the entire K-16 curriculum, if the outcomes are to be achieved. In this respect, the recommendations of national standards committees to include inquiry and critical thinking at all levels of the curriculum are very well motivated. This need for deep and wide curricular penetration underscores the need for research and development. We do not yet know how to achieve it in all the fields, at all the levels, and in all the settings where it is needed. The challenges in the undergraduate curriculum alone are formidable, given the disciplinary content goals that we have for majors and the small amount of time that non-majors spend studying science.

RELATION OF HIGHER-ORDER OUTCOMES TO COVERAGE AND PREREQUISITE KNOWLEDGE

Traditionally, the primary learning objective of college science instruction has been the mastery of disciplinary content. Although faculty members are generally ready to embrace higher-order objectives as well, they are often less willing to experiment with novel modes of instruction and assessment because of a reluctance to sacrifice content coverage and a belief that the pursuit of higher-order outcomes should be delayed until the student has acquired some prerequisite knowledge.

In a straightforward sense, the sacrifice of coverage is not a myth. Instructors who pursue higher-order outcomes seriously usually make changes in the classroom that lead to a reduction in the amount of material that is presented in the classroom. Instructors who choose to lecture, for example, tend to spend time explaining how a particular concept is tied into other concepts in the course, how it was established empirically, why alternatives were rejected, what the current research issues are, and so on, rather than just defining the concept, giving an example, and moving on. In our classroom observations, we have found that the simple presentation of content is significantly reduced in inquiry-oriented classrooms.

The reduction of coverage may be less significant than it appears at first sight, however. Students do not necessarily master concepts just because they are presented in class, and memory-oriented assessments often do not probe the degree of mastery. There is substantial evidence from cognitive psychology that memory is fragile for concepts that are not processed meaningfully or elaborated richly with respect to other knowledge. There is evidence from studies of learning in physics [16] and other disciplines that students sometimes fail to grasp the meanings of concepts even though they can pass memory-oriented assessments and solve familiar types of problems algorithmically. Further, time spent on core meanings and relationships in class may facilitate learning outside of class. If coverage is measured in terms of long-term concept mastery by students rather than mention in the classroom, then it must be said that we do not really know whether, when, and in what settings a commitment to higher-order outcomes necessitates a sacrifice in coverage.

Similarly, there is now abundant evidence from both cognitive psychology and educational research that, in appropriate learning environments, students at any level can engage in scientific inquiry, reason with concepts, and make metacognitive gains [17]. The failure to infuse higher-order cognitive goals into the entire K-16 curriculum potentially has three negative outcomes. First, students may not achieve stable higher-order outcomes during the course of their schooling because they have not spent enough time developing them. Second, student learning of field-specific concepts may fail to be persistent and cumulative, because it is too oriented toward memorization for relatively shallow assessments. Third, students may develop beliefs about the nature of knowledge and learning that lead them to resist instruction that is oriented toward inquiry and critical thinking when it is offered.

An approach to research and development for higher-order outcomes

The preceding discussion supports a research-oriented approach to the continued development of instructional regimes that favor higher-order cognitive outcomes. Theoretical and empirical results in cognitive science and a growing body of educational research support the claims that higher-order outcomes are achievable and that they contribute to the ability to remember and apply field-specific knowledge [1]. The same bodies of research indicate that higher-order outcomes are complex and difficult to achieve. They should be treated as lifetime learning goals that are a continuous part of all K-16 learning environments. Building and maintaining effective learning environments requires continuous interaction between the more theoretical and the more applied aspects of the science of learning. We are not now, nor should we expect ever to be, at a point where a theory of learning coupled with a fixed body of educational research simply deliver final conclusions for immediate implementation in the classroom. There is no more reason to expect the science of learning to be permanently settled than there is to expect final closure in any other science, and because applied science is always more than a simple reading of some more general theory. Significantly, the unfolding of the sciences of learning depends in part on the extent to which education itself is a source for important questions, evidence, and theoretical insights.

An implication of this position is that teachers should think of themselves as applied cognitive scientists and of their classrooms as laboratories for the study of learning. By all means, they should try to implement some version of current "best practices" or "what works" in their classrooms, but, more importantly, they should be involved in some form of research, from confirming what in fact does work in their classrooms [2] to more ambitious attempts to contribute to the evolving sciences of learning and instruction. The potential for such contributions is enhanced when classrooms are sites for research that involves collaborations among educators, educational researchers, and cognitive scientists.

A model for sustainable change: Classroom-oriented research partnerships

We implemented the above approach by founding a research center with a defined agenda that enters into partnerships with faculty groups that involved in educational innovation and assessment. The Center for Research in Education and Learning (REAL) at Hampshire is based on the following objectives:

- To perform research that contributes to systematic knowledge about learning, teaching, cognitive development, and effective uses of emerging educational technologies. Our current focus is on higher-order cognitive outcomes in science education. The goal here is go beyond a "program evaluation" approach that often produces no contribution to the research literature or to national dialogs on learning and instruction.

- To form partnerships with naturally-occurring reform efforts, characterized by sustainable institutional and faculty commitment. The pre-existing interests and work of the faculty groups guarantee a genuine intellectual dialog with our research staff and institutionalized programmatic results will persist beyond particular research projects and outside funding.

- To describe existing, sustainable reform-oriented learning environments in detail. This descriptive research maximizes the chance of discovering phenomena that are of value to the learning sciences and to theories of instructional design, and it provides a framework and baseline for feasible redesign and reassessment of the environments.

- To undertake rigorous assessments of student outcomes. Research-oriented assessment offers theoretical frameworks and levels of confidence that are not possible in ordinary student assessment. It can therefore provide strong direction for both the further development of learning environments and the redesign of standard assessment practices.

- To use project research to refine and redesign the learning environments being studied. The most immediate value of the research should be in the insights it provides into how to change the environment being studied. The resulting cycle of change should lead to new theoretical insights and instructional practices that are worthy of dissemination.

This partnership model has the potential to support sustainable change, because it works at the intersection of careful, theory-driven research and feasible educational practice.

Research methods

The partnership model is consistent with a wide range of research methods, and our methods have changed and will continue to change over time. However, the methodological regime that we have employed for our most highly-developed partnerships over the last five years of work has some notable features that are worth describing.

COMPARATIVE METHODS

First, we have felt that it is important to have some research methods that can be applied comparatively across instructional settings. The point is not to declare winners and losers but rather to gather knowledge about variations in educational goals and practices and their effects on student outcomes. We have employed three complementary comparative methods:

- *Inquiry-oriented classroom observation protocol*. We developed a method of classroom observation that allows a trained observer to code a classroom session for inquiry-oriented content. This instrument has allowed us to compare classrooms on the types of inquiry content that are emphasized. For example, we distinguish among hypothesis generation, research design, and the critique of empirical results. Methods of instruction are coded independently of inquiry content, which prevents their confounding. For example, some lecturers spend quite a bit of time on inquiry, and some highly interactive classroom discussions do not have high levels of inquiry content.

- *Essay-style survey of scientific thinking [STS]*. We developed a set of essay questions that present simple scientific scenarios and require several types of scientific reasoning to answer. The scoring rubric allows the questions to be answered successfully with no use of disciplinary knowledge, and the questions do not engage field-specific knowledge that most undergraduates possess. For example, one question concerns an experiment that investigates corn yields for differing hybrids given differing amounts of water. The question engages general reasoning processes about experimental design and has never yielded answers that contain technical information about farming. Because of its lack of disciplinary content the STS can be administered in any classroom without being affected by the details of the subject matter being taught.

- *Intensive epistemological interviews*. We have employed two versions of an intensive face-to-face interview that probes students' views of the nature of scientific knowledge and inquiry. The questions do not concern the subject matter the students are currently studying and thus can be administered across contexts. The questions deal with the student's understanding of the role of theories, evidence, and uncertainty in science.

SITE-SPECIFIC METHODS

We have found that the full understanding and refinement of a learning environment also requires that specific methods be developed in collaboration with the partner faculty.

- *Intensive case studies*. Our standardized classroom observation protocol yields a fairly coarse-grained analysis of classroom process. It is useful to follow up on these observations with intensive studies of video or audiotaped classroom or simulated classroom sessions to explore in detail the ways in which instructors and students actually engage in scientific reasoning processes.

- *Targeted assessment*. Similarly, the STS can not get at inquiry-oriented learning goals that involve disciplinary content or at progress toward more general higher-order skills that does not yet transfer to the spare and rather arbitrary items that are employed on the STS. We work with collaborating faculty to design assessments that tap progress on scientific thinking while staying closer to the disciplinary context of the course. These research items then serve as prototypes for redesigning assessment practices in the course.

- *Design studies*. Detailed study of classroom process and targeted assessment can begin to interact with each other in a cycle that creates a design study in which a course is progressively altered within the space of an academic term.

There is considerable synergy between comparative and site-specific collaboratively-developed methods. Comparative methods can strongly confirm the basic characteristics of a learning environment and help those who developed it to set an agenda for future development. Methods developed or adapted to study a particular environment strengthen research conclusions and deliver the information that can actually guide the further change. Generally, the use of multiple, complementary methods is a critical strategy in educational research, where single studies rarely deliver definitive conclusions.

An example

THE SETTING

For two years one of the REAL research partnerships has involved the introductory biology staff at a large university. Introductory biology instruction is delivered in this department in large lecture sections of 200-400 students. Currently there are no small recitation sections. Machine-scored multiple-choice examinations are the primary mode of student assessment. The department, led by several faculty members, has made several strong efforts to improve the quality of instruction without changing class size, which is unlikely to change in the near future.

- *Laboratories*. The teaching assistants in the course lead a weekly series of laboratories, which are taught in sections of about 35 students, and which were redesigned several years ago. Several of the labs are structured to engage students in inquiry.

- *Web-based course support*. The department developed its own world-wide-web-based course support system. In the introductory course the system is used to deliver information related to the topics of the class meetings and to log students answers to quiz items that are made available prior to each class meeting.

- *Classtalk*. The three primary instructors for the introductory course dropped traditional lecturing and adopted a small-group discussion and report model for the class meetings. In this model the instructor poses a questions to the class, which are discussed by small groups of students. The groups then report their results to the whole class and the instructor manages a discussion in which groups explain and defend their reasoning. The instructors design questions that engage students in reasoning about scientific concepts but are suitable for brief discussion. When possible the classes meet in lecture halls equipped with Classtalk technology [18]. Classtalk allows small groups of students to post their answers to multiple-choice questions to a central computer, which displays a histogram of the choices on a screen at the front of the lecture hall.[5] The spread of answers on the histogram is an effective stimulus to the whole-class discussion that follows the small-group discussion. In our biology study 2-4 discussion cycles occur during each class meeting.

The lead faculty members received a private-foundation grant to implement and evaluate the web and lecture-based instructional reforms. Our partnership involves these faculty members, the university's institutional research office, and the REAL staff. Research funding has come from the biology department, the private-foundation grant, and REAL's NSF funding.

INITIAL RESEARCH

The collaboration began with an assessment of the intructor's goals for the reform sections of the introductory course. Their goals, which were in part institutionalized in a formal departmental statement, aligned very well with the list of higher-order outcomes sketched at the beginning of this paper.

Hypotheses

Initially, our research focused on two hypotheses:

- Students in the reform sections make significant gains in scientific thinking skills during one semester of instruction. Potentially, gains could also be measured relative to comparison courses, as well.

- Instruction in the reform sections contains significant content related to specific higher-order outcomes. Potentially, content could be shown to be higher than that in comparison courses.

Methods
In the initial semester of research on the course there were two reform sections (each with over 200 students) and a comparison section (also with over 200 students) that used traditional lectures rather than Classtalk. We emphasized comparative methods in this initial round of research:

- The STS was administered pre and post semester to most students in the two reform and the comparison classes. Trained readers, who were blind to section membership and pre-post status, scored the surveys using a standard STS rubric.

- An experimental multiple-choice version of the STS was administered to a sample of students from the three classes.

- Observers coded representative class meetings of all three sections.

- Five students from each section were given the intensive epistemological interview. These interviews were part of a larger, ongoing comparative project and are not discussed further here.

Results
We found that the reform sections did include significant content related to higher-order outcomes, both in absolute terms and relative to the comparison section. Most of this content was delivered during Classtalk discussion cycles. Reasoning with causal models was by far the cognitive skill that was most emphasized among the various higher-order goals. Understanding experimental designs and graphical interpretation were also significantly represented.

Although students in small-college inquiry-based classes have consistently shown gains on the essay-style STS, neither the reform nor the comparison university biology sections made pre-post gains. The reform sections did make significant pre-post gains on the experimental multiple-choice version of the STS.

Our interpretation of the failure to see gains on the STS was that it emphasizes reasoning skills that were not heavily represented in our classroom observations and that it requires students to explain their reasoning in writing, a skill that was barely represented at all in the reform sections. The gains on the multiple-choice version of the STS tend to confirm this interpretation. The multiple-choice items did not require explaining one's reasoning in writing and were actually similar in format to the Classtalk items that the reform students had extensive experience with. Although multiple-choice items make it difficult to score the details of a student's reasoning,[6] they may have provided a more sensitive indicator of modest student gains in this case by removing the necessity for writing and by the tendency of the alternative answer choices to scaffold students' reasoning processes.

TARGETED RESEARCH

The initial research results led the research partners to several provisional conclusions:

- The learning goals of the course should be reviewed, and the instructors should focus on those goals during class time.

- Student assessment in the course should place more emphasis on the higher-order learning goals to make it clear that there is a relationship between work toward those goals and course grade.

- An assessment instrument should be designed that taps the higher-order thinking skills emphasized in the course using questions with college-level biological subject matter.

- The character of student reasoning during Classtalk discussions should be studied in more detail to determine whether students are practicing higher-order skills.

The reasons for these conclusions are fairly self-evident. The relatively modest findings of the initial research suggested that the treatment should be intensified and the measurement instruments be made more sensitive.

The review of course learning goals revealed that the instructors were particularly committed to helping students learn to reason qualitatively with causal models in cellular and molecular biology, to make logically sound inferences from experimental data, and to move back and forth between concepts and graphical presentations of data. The instructors and researchers worked together to design a new instrument that assessed these skills directly in biological contexts and which could be administered as a course assignment. One of the instructors also decided to offer a voluntary once-a-week Classtalk-style discussion section in a large seminar room with moveable tables. This arrangement allowed for high-quality videotapes to be made of student discussions.

Hypotheses

- Students in the reform sections make significant gains during the course in targeted reasoning skills assessed in the context of biology. No comparison section was available for this phase of the research.

- Students practice and improve in the targeted reasoning skills during Classtalk discussions.

Methods

- Pre and post assessments that targeted course-specific higher-order cognitive outcomes were administered. Most items consisted of a stem portion that presented a scientific situation and a cumulative series of multiple-choice items, each of which asked students to defend their choice in a short written answer [See Appendix A for an example]. One question was repeated on the pre and post tests, others were designed to be parallel. Only the repeated question, on mitochondria [Appendix A] concerned material that was taught directly in the course.

- Classtalk groups were videotaped in the small discussion section throughout the term. The analysis of the tapes is ongoing, and they will not be discussed further in this paper.

- We observed the class to confirm that the implementation of the target content in the Classtalk cycles.

Results

The targeted assessment turned out to be sensitive enough to measure student change. Students improved in their ability to reason with a model that was taught during the course (electron transport across the inner membrane of the mitochondrion) and in their ability to reason with models they had not studied. The latter conclusion rests on the comparison of answers to the mitochondrion question on the pre-test and to a question about the mechanisms involving cyclic AMP on the post-test. Students' ability to assess the logic of inference from the data for a given experimental design also improved. Students did not improve in their ability to sketch graphs or to draw appropriate inferences from graphs displaying correlational data.

These assessment findings accorded well with classroom observations, which showed that the instructors devoted much more time to reasoning with models and to design-related inferences from data than they did to graph interpretation and the correlation vs. causation issue.

It is important to note that although the targeted assessment items all had biological content, the results indicate both improvement in key higher-order cognitive skills and transfer. With the exception of the mitochondrion question, none of the subject matter on the post-test had been covered in the course.

CONCLUSIONS

The example research program has demonstrated so far that it is possible to persistently engage students with some aspects of scientific reasoning in a large university lecture course in biology. That investment has had a measurable impact on students' reasoning skills, which showed transfer from the course material to other topics in biology. The impact seems to differ from and to some extent fall short of what is achieved in inquiry-oriented settings characterized by small class sizes and frequent writing assignments. The outcome seems to have been achieved with little or no sacrifice in subject-matter coverage, although this conclusion has not been fully substantiated through a study of course records.

Ongoing research will contribute to this picture. Specifically, we are still analyzing the intensive epistemological interviews and the videotapes of Classtalk discussions. In a further research cycle we plan to undertake an analysis of the effects of student characteristics on higher-order outcomes, to do further assessments to explore the utility of multiple-choice questions and the effects of further enhancements to the course, and to further incorporate inquiry-oriented assessment into the course examinations.

Discussion

Our view is that we can not make significant and sustainable progress in achieving higher-order cognitive outcomes in science without the kind of detailed collaboration between researchers and disciplinary scholar/teachers illustrated above. Although higher-order outcomes are to a greater or lesser extent independent of specific disciplinary content, they will be largely achieved within the context of instruction in particular disciplines. We only confront the challenges involved if we work in authentic disciplinary settings.

References

[1] J. D. Bransford et al. (Eds.), *How people learn: Brain, mind, experience, and school, expanded edition,* National Academy Press, 2000

[2] K. P. Cross & M. H. Steadman, *Classroom Research: Implementing the scholarship of teaching*, Jossey-Bass Publishers, 1996

[3] S. Dahaene, *The Number Sense: How Mathematical Knowledge Is Embedded in Our Brain,* Oxford University Press, 1997

[4] G. F. Marcus, *The algebraic mind: Integrating connectionism and cognitive science,* MIT Press, 2001

[5] J. L. McClelland, D. E. Rumelhart, G. E. Hinton, The appeal of parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, and the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations,* MIT Press, 1986

[6] P. Smolensky, On the proper treatment of connectionism, *Behavioral and Brain Sciences, 11,* 1-74, 1988

[7] C. Lebière and J. R. Anderson, A connectionist implementation of the ACT-R production system, *Proceedings of the fifteenth annual conference of the Cognitive Science Society,* 635-640, Lawrence Erlbaum, 1993

[8] P. W. Cheng and K. J. Holyoak, Pragmatic reasoning schemas, *Cognitive Psychology, 17,* 391-416, 1985

[9] G. T. Fong and R. E. Nisbett, Immediate and delayed transfer of training effects in statistical reasoning, *Journal of Experimental Psychology, 120,* 34-45, 1991

[10] D. Kuhn, *The skills of argument,* Cambridge University Press, 1991

[11] Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

[12] Nisbett, R. E. (Ed.). (1993). *Rules for Reasoning*. Hillsdale, NJ: Erlbuam.

[13] Flavell, J. H., Miller, P. H., & Miller, S. A. (2001). *Cognitive Development, 4th Ed.* Prentice-Hall.

[14] King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey-Bass.

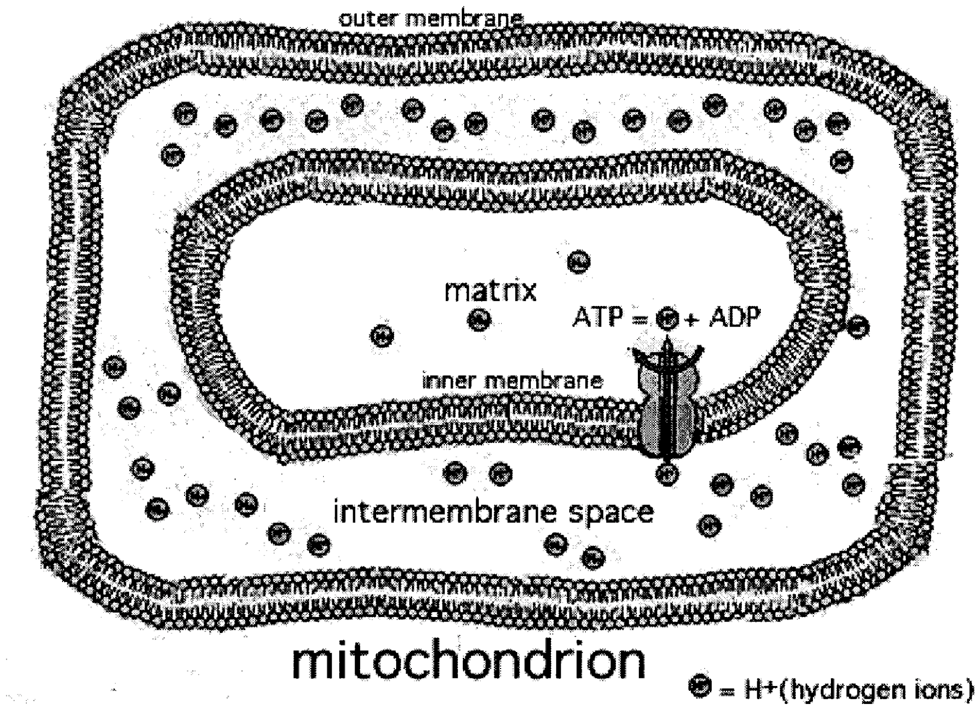
[15] Perry, W. (1970). *Forms of intellectual and ethical development in the college years*. New York: Holt, Rinehart, & Winston.

[16] L.C. McDermott and E.F. Redish (1999). "Resource letter on Physics Education Research," Am. J. Phys. 67 (9) 755

[17] Carey, S. & Smith, C. (1993). On understanding the nature of scientific knowledge. *Educational Psychologist, 28*, 235-51.

[18] 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.

Appendix A: An example hybrid multiple-choice-essay question



In the mitochondrion, the electron transport system pumps protons (H^+) from the matrix into the intermembrane space. This creates a gradient of H^+ (more in the intermembrane space than in the matrix). Both membranes are impermeable to H^+ (protons can not cross the membrane with out help from enzymes). ATP synthase is an enzyme complex in the inner membrane that uses the flow of H^+ down the gradient (back into the matrix) to produce ATP.

2a. If you could increase the concentration of H^+ in the intermembrane space it would

A. stop the production of ATP immediately
B. increase production of ATP.
C. decrease production of ATP.
D. have no effect on the production of ATP.

Why?

Continued on next page

2b. If you could increase the concentration of H^+ in the matrix to exceed the concentration in the intermembrane space it would

A. stop the production of ATP immediately
B. increase production of ATP.
C. decrease production of ATP.
D. have no effect on the production of ATP.

Why?

2c. If you could add a detergent that would make the outer membrane permeable (H^+ could cross the outer membrane and leave the mitochondrion without help from enzymes), it would

A. stop the production of ATP immediately
B. increase production of ATP.
C. decrease production of ATP.
D. have no effect on the production of ATP.

Why?

[1] The Center for Research in Education and Learning (REAL) was founded in 1997 to study and enhance learning and intellectual development in college and pre-college students. REAL's mission is to conduct high-quality research that promotes the refinement, credibility, and acceptance of successful educational innovations. The initiative, based at Hampshire College, is cross-institutional, and its research team comes from several colleges and universities. REAL draws on Hampshire's long experience with inquiry-oriented and interdisciplinary education, as well as on its unique strengths in cognitive science, and technology-based (as well as traditional) educational materials development. REAL's current projects concern instructional practices and learning outcomes in inquiry-oriented science education. In addition to the author REAL's research staff, graduate students, and advisors over the last five years have included Mary Anne Ramirez, Laura Wenk, Carol Smith, Tom Murray, Loel Tronsky, Samia Khan, Alisa Izumi, and John Clement. Collaborating departments have included the School of Natural Science at Hampshire College, the departments of biology and chemistry at the University of Massachusetts, and the department of biology, the department of earth & environment, and the Unity of Science Program at Mount Holyoke College. Funding comes from the NSF and other agencies.

[2] The student outcomes that are of interest in this paper are spelled out in some detail later in the paper. The use of the term *higher-order* to refer to these outcomes is somewhat nonstandard and is a matter of convenience in this paper.

[3] There is now extensive evidence that mathematics is supported by distinct regions of the brain [3].

[4] An abstract cognitive structure can be understood as equivalent to a symbolic expression that contains variables, into which specific values from an appropriate, possibly infinite, set can be substituted. Algebraic expressions, such as $y = mx + b$, are classic examples. Abstraction away from specific detail is endemic in human cognition, since it not only allows us to learn obviously abstract material, such as mathematics, but also to recognize dogs or chairs of differing appearance and even our own mothers on different occasions. How abstract concepts are represented and learned is a subject of intensive research and significant theoretical ferment in cognitive science. Some theorists, such as Marcus [4] and Anderson [7], have proposed that symbolic expressions with variables and processes for variable instantiation (or binding) are directly represented in the brain. Others, such as McClelland, Rumelhart, & Hinton [5], and Smolensky [6], have proposed that in at least some cases the necessary kinds of generalization and transfer are achieved by parallel, distributed neural networks that are driven only by the statistical similarity among particular examples. Empirically-oriented education researchers might be drawn to symbolic or hybrid symbolic-similarity positions, which suggest experimentation with a wider range of training regimes and learning environments.

[5] Classtalk systems require that each student or small group of students be able to transmit a multiple-choice answer electronically to a central computer that can plot and display a histogram of the answers. The communication channel and software can also allow student log-in, which enables various forms of record keeping, such as lecture attendance. The study classrooms were wired with outlets that allowed students to plug in their graphing calculators and log into the class meeting and record their answers. Wireless or standard ethernet-based technologies could also be used to implement Classtalk.

[6] Each item on the multiple-choice version of the STS included a correct alternative that represented a sound line of scientific reasoning and incorrect distractors that were developed from common flawed responses to the essay version of the assessment. The intention in developing this type of item is that the researcher can infer the student's reasoning from the alternative selected. The validity of the inference depends very much on the care with which the instrument is developed and validated. The M-C STS had undergone no statistical validation prior to this research. Ultimately, if one is interested in the student's ability to generate lines of reasoning, it is impossible to equate essay and multiple-choice items, because the multiple-choice alternatives either directly supply or cue reasoning processes that students must generate in writing essays.