A Faculty-Development Model for Transforming Introductory Biology and Ecology Courses

CHARLENE D'AVANZO, CHARLES W. ANDERSON, LAUREL M. HARTLEY, AND NANCY PELAEZ

The Diagnostic Question Cluster (DQC) project integrates education research and faculty development to articulate a model for the effective transformation of introductory biology and ecology teaching. Over three years, faculty members from a wide range of institutions used active teaching and DQCs, a type of concept inventory, as pre- and posttests to assess students' understanding of concepts about energy and matter across biological scales of organization. Surveys of the instructors indicated a substantial use of DQCs and active teaching, and nearly all of those faculty members participating in the research saw significant student gains and a large positive effect size between the pre- and posttests. Important programmatic components included reliable research-based conceptual questions and the associated active-learning exercises; formative examination of preinstruction data, including the students' written answers; a professional society for recruitment, workshops, and dissemination; progressive faculty growth over three years; and cooperative communities of practice. We propose that research-based conceptual inventories can be effective tools in faculty-development programs offered through biology professional societies.

Keywords: faculty development, introductory biology, diagnostic question clusters, diagnostic questions, misconceptions in biology and ecology

Despite considerable effort to improve biology teaching, many faculty members acknowledge that students, especially those in introductory courses, routinely cannot apply core concepts and logical reasoning to fundamental questions—that is, they do not "think like biologists" (e.g., D'Avanzo 2008, Momsen et al. 2010, Smith JI and Tanner 2010). The recent publication *Vision and Change: A Call To Action* (AAAS 2011) advocates that to promote such thinking, faculty members need to shift focus "away from presenting all the facts (i.e., 'covering the material') and toward clearly articulating expected student outcomes and following students' progress in achieving these outcomes" (p. 22). Because calls for biology instructors to limit course content and to focus on clear outcomes and conceptual understanding are far from new (NRC 2003, 2009), why is this recent report asking for similar reforms? We propose that this is partly because such a fundamental transformation of biology instruction necessitates considerably more time and effort and better pedagogical tools than many understand; a major shift in a faculty member's beliefs about teaching biology; effective rewards for faculty members who make the effort; effective learning communities for faculty members striving to reform their teaching; and research-based, explicit faculty-development models that work for the wide range of faculty members teaching these courses. The complex interplay of these components helps us understand why progress on significant transformation of biology teaching, particularly at the introductory level, has been fairly limited (e.g., Ebert-May et al. 2011).

In this article, we address a faculty-development program in which faculty members use research-based diagnostic question clusters (DQCs; table 1), a type of concept inventory, as pre- and posttests to examine students' understanding of biology concepts and use active teaching to help students improve. These DQCs assess the students' ability to apply the principles of conservation of matter and energy across a biological scale, and they are focused on commonly taught topics. Parallel pre- and posttests examine similar processes and reasoning. The DQCs are used to identify problematic reasoning (e.g., matter can turn into energy in a biological context) and are therefore *diagnostic*; they are *clustered* to identify patterns in regard to ability, scale, and processes (e.g., the ability to trace matter during photosynthesis in cells and ecosystems; tables 1 and 2). The DQCs, framework, active-teaching exercises, and supporting information are on the open-source "Thinking Like a Biologist" Web site (*http://biodqc.org*) and are explained in Wilson and colleagues (2006) and in Hartley and colleagues (2011). The present article complements the work of Hartley and colleagues (2011), which was focused on student performance.

Diagnostic tests, such as concept inventories, are promising tools for biology faculty-development programs because questions validated by research help the faculty members

BioScience 62: 416–427. ISSN 0006-3568, electronic ISSN 1525-3244. doi:10.1525/bio.2012.62.4.12

Table 1. Description of the energy and matter diagnostic question clusters (DQCs), which are based on a framework concerning the application of the principles of energy and matter conservation across biological scales of organization.

Focus and characteristic	Description
Principles	Conservation of matter and energy
Scales	Subcellular–cellular–organism–ecosystem
Topics	Photosynthesis, respiration, oxidation, digestion, growth, movement, transformation of organic and inorganic molecules, energy flow, decomposition
Diagnose	Designed to expose students' naive conceptions and understanding of energy and matter across biological scales of organization; incorrect answers are "distractors" identified by research
Clustered	Sets of eight questions focusing on problematic reasoning tendencies about a topic
Type of question	Application to novel situation
Format	Multiple choice, multiple true–false, and open format when students choose an answer then briefly explain
Timing	Completed by most introductory level students in 15 minutes
Pre- and postinstruction questions	Matched sets of questions are used before and after instruction
Pretest	Provide formative data on students' alternative conceptions and poor understanding prior to a module or course
Posttest	Provide summative data on possible improvement in students' understanding at the end of a module or course
Scoring	Faculty score extended responses as principled (scientific), mixed, or informal (naive) reasoning; answers sampled in large classes
Student-active teaching	DQCs are linked to specific student-active exercises designed to improve diagnosed problems

Note: The DQCs are on the "Thinking Like a Biologist" Web page (*http://biodqc.org*).

focus on key concepts and provide reliable data on students' conceptual comprehension, and therefore on the effectiveness of their teaching (Garvin-Doxas et al. 2007, D'Avanzo 2008, Michael et al. 2008). For instance, Marbach-Ad and colleagues (2009a) found the host–pathogens interactions (HPI) diagnostic test essential to the curricular reform and professional development of nine linked microbiology courses. Examination of HPI pre- and posttest data by the 19 faculty participants "anchored and deepened" (Marbach-Ad et al. 2009b, p. 408) discussions about student learning, leading to more-effective reforms. This team fits Rogan and Anderson's (2011) definition of a faculty-development learning community—professors working collaboratively toward course reform, bound together socially and professionally. Learning communities employing concept inventories may thus be powerful vehicles for the transformation of college biology instruction.

For several years, we have administered an integrated education research-faculty-development program in which faculty members use DQCs linked to active teaching exercises. Both the question sets and the associated active exercises are designed to improve students' understanding of fundamental content and concepts regularly taught in introductory biology and ecology courses: processes related to energy and matter across biological scales of organization. Numerous reports have shown that student-active teaching can help students learn biology concepts more effectively (e.g., Knight and Wood 2005), but few studies have been focused on faculty development or what it takes for faculty members to transform their own courses with active teaching. In an earlier article, Hartley and colleagues (2011) reported on students' pre- to posttest learning gains in courses taught by the DQC participants. In contrast, faculty members are the focus of the present article. Our aim is to articulate important components of faculty-development programs that will lead to the transformation of biology teaching and learning across a range of institutions. To do so, we describe the DQC program and key findings with individual faculty participants as the unit for analysis.

A main goal of the DQC project is to help faculty members focus on their students' conceptual change, but what does this mean? We rely on Tanner and Allen's (2005) definition of *conceptual change* as "a learning process in which an existing conception ([an] idea or belief about how the world works)… held by a student is shifted… away from an alternative or misconception and toward the dominant conception held by experts in the field" (p. 113). Therefore, a first step in teaching for such understanding is to recognize students' naive ideas about concepts and ideas key to a course. Alternative conceptions have been important in biology education research and have resulted in extensive lists of misconceptions (e.g., Michael et al. 1999, Anderson DL et al. 2002). Although they are useful, such lists may not help faculty members realize why students fundamentally hold these ideas and thus to realize how to design courses that address those ideas. We build on prior work to identify problematic patterns in students' reasoning that span the typical content of introductory courses and that generally limit reasoning about biological phenomena (Hartley et al. 2011). Our research began with interviews of university students, in which we studied incorrect responses to multiple-choice questions concerning energy and matter (Wilson et al. 2006). These interviews showed that these students did not use reasoning that was based on the laws of conservation of energy and matter and had great difficulty applying what they had learned at one scale of biological organization

*Table 2. Example diagnostic question clusters (DQCs) question and written student responses to the first of eight questions in the Forest Carbon DQC (*http://biodqc.org*).*

			Percentage of the answers as a function of reasoning type		
DQC question	**Example student response**	**Type of reasoning exhibited**	**Principled**	**Mixed**	**Informal**
Explain your ideas about how plants, animals, and soil in a forest interact with carbon dioxide.	All organisms produce CO_2 through cellular respiration and all store carbon for growth. Plants also take in CO_2 during photosynthesis, which produces O_2 plus glucose for the plant itself.	Principled			
Which groups of organisms absorb carbon dioxide from the atmosphere? (Circle all correct) Explain your answer. *Plants* Animals Soil Organisms	All living organisms including plants, animals and microorganisms are composed of carbon.		14	16	20
Which groups of organisms release carbon dioxide into the atmosphere? (Circle all correct) Explain your answer. *Plants Animals Soil Organisms*	Plants absorb carbon dioxide and produce oxygen. Animals do the reverse.		26	71	0
	Plants absorb carbon dioxide and change it for oxygen.				
	Plants release oxygen for animals to take in.	Informal			
	Animals pass on carbon dioxide by defecation and urination.				
Which groups of organisms store carbon? (Circle all correct) Explain your answer. *Plants Animals Soil Organisms*	Soil is responsible for absorbing carbon dioxide so plants and animals can use it.		31	42	22

Note: The correct multiple-choice answers are in italics. Distractor responses were validated by student interviews (Wilson et al. 2006). The processes tested are cellular respiration, photosynthesis, and the transformation of carbon in organic and inorganic forms; the principle is tracing matter; the scale is the organism. The percentage values are postinstruction and exclude those students who did not answer the question ($N = 174$).

(e.g., cellular) to another (organism or ecosystem). Because of this, the resulting research-based DQCs are organized within a framework based on the conservation of matter, the conservation of energy, and scales of organization (table 1; Hartley et al. 2011).

Specifically, in this article, we examine what it takes for faculty members to design introductory courses around both a conceptual understanding of one of the *Vision and Change* (AAAS 2011) key concepts and clear outcomes and evidence that students are learning identified concepts. To do so, we examine the reactions of the faculty participants to DQCs, active learning exercises, and the overall program, and we interrelate program components reported by the instructors whose students exhibited large pre- to posttest learning gains in an attempt to identify essential program elements.

The project design included careful attention to faculty participant selection, including their recruitment process and teaching experience; intervention of both content (e.g., energy and matter DQCs) and process (e.g., dissemination at meetings); and multiple measures of outcomes and impacts (table 3). In such studies, the triangulation of data from several sources leads to a more comprehensive picture of the participants' experience (Cohen L et al. 2000).

Faculty-development program

Since introductory biology and ecology courses are widely taught, we worked with equal numbers of faculty members from universities, four-year colleges, and community colleges.

Faculty and courses. The participating faculty members represented five large (with more than 10,000 students) and eight medium-size (2000–10,000 students) universities, as well as two small colleges (with fewer than 2000 students). One institution was a historically black university, and three others had minority populations larger than 30%. The faculty members were a roughly equal mix of assistant and associate professors, who, on average, devoted 68% of their time to teaching (a range of 40%–98%). Each of the 15 faculty members selected one course for this program; seven of the courses were introductory biology, seven were introductory ecology or environmental science, and one was an intermediate-level ecology course. About a quarter of the courses were cotaught.

Process. The faculty members were recruited through an electronic mailing list and announcements sponsored by the Ecological Society of America (ESA). We chose the ESA for

Table 3. Key components (expressed as inputs) and outcomes (results) of the Diagnostic Question Cluster (DQC) project.

Inputs	Outcomes
People: Faculty members selected through application process are experienced with active teaching; the faculty members teach introductory biology or ecology at universities, four-year colleges and community colleges; the program organizers are leaders in faculty development and education research; the participant instructors and program leaders are seen as colleagues	Participants from wide range of institutions have similar experiences and outcomes Majority of faculty members see significant pre- to postinstruction DQC gains, but many students lack basic understanding at the end of the course Program organizers and faculty participants work collaboratively to improve the project
Process: Three-year faculty-development program with annual workshops; the faculty members are recruited through the Ecological Society of America (ESA) and work in teams; the instructors score the students' answers and contribute to an optional research study; the instructors disseminate their work at venues they value; the program provides tools (DQCs and active-teaching approaches) and support (e.g., help coding extended responses).	Trust of validated DQCs give the instructors confidence Expanding the formative use of DQCs by the instructors over time The faculty members become more intentional about their teaching practices The majority of the instructors contribute to the project research The faculty members drive the workshop agenda in years two and three Very active dissemination at the ESA meeting, regional meetings, home institution; community-college faculty members most value regional venues Self-selected teams are the most productive with dissemination
Content: The DQCs are focused on the conceptual understanding of energy and matter transformations or pathways; the active-learning targets diagnose problems in understanding or reasoning and stress cooperative learning; the DQC framework and questions clearly define learning outcomes; the DQCs are data driven; the robust, practical, open-source Web site (*http://biodqc.org*) includes DQCs, active-learning approaches and an introduction on their use.	The instructors use DQCs, scoring important to this process DQCs are improved by the participants The instructors use targeted active-teaching exercises The instructors develop new active-teaching exercises The instructors contribute to Web-site improvement

recruitment and for workshops and dissemination because numerous reports (PKAL 1991, AAAS 2011) identify professional societies as important for faculty professional development. In 2008, 15 faculty members were chosen through written applications; to participate, they had to demonstrate experience with student-active teaching, show a clear understanding of project goals, and make a three-year commitment.

There were three annual all-day workshops. The first, in 2008, was focused on the DQC content framework; how questions were developed; how to administer, score, and interpret student data; linking active teaching strategies with specific problems in student understanding and thus the framework; and ideas for use of DQCs and active teaching in upcoming courses. We also described a model for use of the DQCs and active-teaching exercises. The faculty members would give a selected DQC as a pretest at the beginning of the course or module, examine these data to assess common conceptual problems, use student-active teaching exercises targeting the identified problems to help the students improve their understanding of these concepts or reasoning, and give the matching posttest to assess gains. We emphasized that this pattern was a goal that most teachers would need to slowly work toward, depending on their prior teaching experience, class size, and student population, for example. The faculty members took a very active role in workshops in 2009 and 2010 by presenting posters and discussing the specific DQCs and active-teaching strategies that they had used.

Teams from similar institutions worked together in workshops and were encouraged to do so during the academic year. We assigned the teams in year one, but some faculty members formed new working groups. Between workshops, the faculty members and program leaders interacted often about issues such as answer coding, the use of active-learning strategies, and help with data interpretation. They scored student responses in the manner described below and sent the data to program organizers at Michigan State University for archival and analysis. All of the data were submitted for smaller classes, and subsamples were submitted for large ones; the scores were validated independently for every tenth response and rescored when agreement was less than 90% (which was rare). In year one, some faculty members participated in an optional research study on the students' pre- and posttest responses to a common set of DQC questions. As with scientific research, dissemination was an important part of the project, and the faculty members were encouraged to present their findings and experiences.

Content: DQCs and active teaching. Many DQCs ask students to explain their selection. To assess students' reasoning abilities, the faculty members scored the students' written answers according to three categories (table 2; Hartley et al. 2011). First, the students using *principled-based reasoning* applied the principles of the conservation of matter and energy across scales of organization; this reflects expert reasoning and understanding (e.g., Chi et al. 1981). Second, the students who used *informal reasoning* did not attempt to trace matter or energy at all and often relied heavily on nonscientific, informal language (such as writing that energy is "burned up") and ideas in their explanations; this represents novice or naive reasoning or understanding. Some of these students also relied on force-dynamic reasoning in which events are said to happen because they "need to" or are caused by outside forces (Mohan et al. 2009). Finally, the students using *mixed reasoning* may have attempted to trace matter or energy, for example, but also displayed common alternative conceptions. Active learning is important in this project because the students' reasoning and understanding difficulties diagnosed by pretests could improve through active-learning exercises that target these problems, which would then be confirmed in the posttest. In the first workshop, the faculty members worked with student-active exercises that we had developed; there are 12 activities on our Web site. The instructors were also encouraged to use, develop, and share their own approaches if they wished.

Evaluation, quantitative analysis, and exploration of the interviews. The findings were taken from a variety of sources so that the patterns that emerged could be more rigorously validated. We evaluated the effectiveness of the project with examples of instructional materials adapted for each instructor's own course, annual written surveys, phone interviews after one semester and annually using a common set of questions, focus-group discussions after workshops, dissemination output, and with pre- and posttest DQC data provided by the instructors. The interview, survey, and focus-group questions were centered on the faculty members' experience with and response to the DQCs, their use of active-teaching exercises, possible changes to courses, the Web site, workshops, teams, and dissemination. The questions were open ended, and the focus-group discussions and phone interviews were recorded. Using N-Vivo (QSR International, Cambridge, Massachusetts), the recordings were transcribed and categorized by ideas and statements frequently reported by multiple instructors and to allow the faculty members to describe their experiences in their own words.

In addition to examining the pre- and posttest DQC data expressed as percentages, we compared the pre- and posttest DQC data from individual faculty members by effect size (ES), which is widely used in psychology and education studies. Calculated here using J. Cohen's *d* equation (Cohen J 1988), ES is the mean posttest score minus the mean pretest score divided by the pooled standard deviation for the DQC data from one course. To estimate the students' proficiency, we applied item response theory (IRT)–based methods—specifically, the partial-credit model (Masters 1982), which was designed for test items with two or more ordered categories. Effect size is useful in this study because the index describes the magnitude of the difference between the pre- and posttest scores (e.g., if an improvement is large or small) and is independent of sample size (which is important in projects with faculty members from a broad range of institutions; Vaske et al. 2002). The values were grouped into small (with an ES of 0.2–0.5), medium (with an ES of 0.5–0.8), and large (with an ES greater than 0.8) changes, following J. Cohen (1988).

The interview transcripts were examined to explore how faculty with high ES scores talked about the program and to give the teachers a voice. We would learn different lessons, for instance, from a professor teaching a new course than from one who found that DQCs did not complement the course's content. The interviews also highlight specific ways in which the individual faculty members responded to the DQC data and how they attempted to change their teaching as a result of the project.

Results of the surveys

Below, we present the results of the surveys and provide analysis thereof.

Use of the DQCs and the associated active-learning exercises. Responding to a written survey at the end of the initial workshop, most of the faculty members strongly agreed that they intended to use the DQCs to examine their students' understanding of energy and matter in ecological contexts, and the scores for this question were quite similar after the instructors' first semester of teaching (table 4). After three years, there was generally strong agreement that the DQCs did help identify poor understanding about these concepts. Furthermore, the instructors carefully examined the wording of the DQCs, and therefore one outcome was validation of the DQCs and, in some cases, an improvement of the scoring method. The faculty-member scores concerning the use of the active-learning strategies were a mix of strong agreement to general agreement (an overall mean of 4.3 on a 5-point scale) and remained similar over the three years (table 4). The higher active-learning coefficient of variation values indicate a wider range of responses than to the DQCs. The most frequently used active-teaching approaches, indicated in written surveys and interviews, are presented in table 5. Some of the faculty members modified the exercises, and many developed additional active-learning lessons. For example, a faculty member teaching about respiration in a large introductory biology course used a sensor to monitor classroom carbon-dioxide concentration over the class period and projected the data for the students to see and discuss.

In the first year, phone interviews showed that nearly all of the faculty members waited until after the end of the course to examine both the pre- and post-DQC data. The initial workshop took place in August 2008, and the instructors were in their classrooms soon afterward; in the interviews, all but one reported that it was challenging to simply administer the pre- and posttests, and they did not have time to score both data sets until the course's end. The second annual workshop was important, because the discussions of actual pre- and posttest data, as well as the responses of several faculty members who were more experienced with using formative data, helped the instructors better understand how to use the pre-DQC data in their own classrooms. By the final year of the project, most of the faculty members reported that they were using pre-DQC data for formative assessment of their students' understanding about matter and energy; those with very large classes examined random subsamples (about 10%) of the pre- and posttest responses.

Student data. The DQCs are critical to faculty development, because the reliable pre- and posttest data help teachers accurately assess their instruction, among other reasons.

Table 4. Abbreviated responses to questions concerning the use of diagnostic question clusters (DQCs) and active teaching at the end of the first workshop, after the first semester teaching with DQCs, and after three years in the project.

Teaching method	Success of teaching method	Timing	Response		Coefficient of variation
			Mean	Standard deviation	
DQCs	Intended to identify difficulties and poor thinking	Right after workshop	4.8	0.1	2.1
	Used to do so	After one semester	4.8	0.3	6.2
	The DQCs were effective	After three years	4.5	0.3	6.2
Active teaching	Intended to be incorporated	Right after workshop	4.4	0.4	9.1
	Materials used to incorporate active-learning strategies	After one semester	4.4	0.7	15.9
	Successfully applied active-learning strategies	After three years	4.1	0.9	21.9

The responses values are on a five-point scale (1, *strongly disagree*; 2, *disagree*; 3, *neutral*; 4, *agree*; 5, *strongly agree*).

*Table 5. Three most commonly used active-teaching exercises on the "Thinking Like A Biologist" Web site (*http://biodqc.org*) as reported by the faculty participants.*

Exercise title	Principles and processes	What students do
Follow the Carbon	Tracing matter at the atomic or molecular, cellular, organismal, and ecosystem scales; generation and transformation of carbon; cellular respiration	Describe in their own words and in drawings what happens to a carbon atom in this scenario: *You are a carbon atom and you are taken in by an oak tree, converted to starch, stored in a seed, which is eaten by a squirrel. In the squirrel, you are stored in fat; the squirrel uses up all of its fat during winter hibernation, which leaves you where?*
Mice in a Box	Tracing energy and matter; making the connection between respiration at cellular and organismal levels	In the context of respiration, they interpret a graph showing rapid increase in temperature in a closed system containing animals (mice).
Dead Cow (Ecosystem Carbon Flow)	Tracing matter, generation, transformation, and oxidation of carbon at cellular, organismal and ecosystem levels	They interpret bar graphs to explain what happens to carbon atoms from a decomposing dead cow passing through microorganisms, plants, and several types of animals over time.

Note: Additional information on the Web site describes how instructors can use the exercises.

After the initial workshop, 13 faculty members contributed to the voluntary research project by using common sets of DQCs focused on energy flow and carbon cycling in their courses. In the context of this research, the majority did see gains in student understanding; the average percentage of students using principle-based reasoning more than doubled from 12% to 27% after instruction (Hartley et al. 2011). However, at the course's end, 50% of the students on average still used mixed reasoning in their responses and 16% exhibited informal reasoning, with no attempt to trace matter or energy. These data underscore the great challenges for both students' development and faculty members' ability to instill in them a genuine understanding of processes such as photosynthesis, respiration, and transformation across biological scales of organization.

Categorized by process, the gains were greatest for photosynthesis questions, and those for DQCs concerning digestion and biosynthesis (Hartley et al. 2011) were small. Categorized by scale, the pre- and postinstruction gains were greatest for questions about ecosystems, moderate for organismal ones, and smallest for questions at the atomic–molecular level (table 6). Common problems include beliefs that atoms could become other atoms, that atoms could disappear, that air "is nothing," that biological phenomena occur because of some "need," and that photosynthesis and respiration are the reverse of each other (table 2; Hartley et al. 2011). Typical student responses for a DQC question show that many of the students could not, for example, correctly identify broad categories of organisms that photosynthesize, respire, or store carbon (table 1). In interviews and workshops, many of the faculty members expressed dismay about some students' poor performance and ignorance of quite basic concepts. However, the pre- to posttest gains also encouraged them. In regard to ES, for 11 faculty members using the same DQC questions, 8 were in the large ES change category, 2 were in the medium range, and 1 had a value of –0.1 (table 7). All positive values indicate significant differences between the pre- and posttests in student responses, according to a paired *t*-test. The professor with the negative value did not teach the proposed

Table 6. Percentage of students applying informal, mixed, and principled reasoning to diagnostic question cluster (DQC) questions concerning matter and energy transformations and pathways at three levels of organization.

Scale	Reasoning type	Pretest	Posttest	Pre- to posttest gain
Atomic or molecular	Informal	26	20	–6
	Mixed	63	65	+2
	Principled	4	12	+8
Organismal	Informal	22	14	–8
	Mixed	57	56	–1
	Principled	10	24	+14
Ecosystem	Informal	21	18	–3
	Mixed	57	42	–15
	Principled	12	34	+22

Note: The data are for students in introductory biology and ecology courses at 13 universities and colleges ($N = 525$; from Hartley et al. 2011).

Table 7. Effect size (ES) data for faculty participants using the same pre- and posttest diagnostic question clusters.

	Number of students				
Faculty member	Pretest	Posttest	ES	ES category	*p*-value
1	42	40	2.46	Large	<.001
2	18	18	1.39	Large	<.001
3	34	34	1.37	Large	<.001
4	47	53	1.24	Large	<.001
5	24	22	1.02	Large	.006
6	33	18	1.00	Large	<.001
7	32	29	0.96	Large	.007
8	20	19	0.94	Large	<.001
9	120	111	0.71	Medium	<.001
10	35	34	0.66	Medium	.008
11	30	29	–0.10	—	.694

Note: The students were randomly sampled in larger classes. ES categories are according to J. Cohen's (1988) definition, described in the text, and the *p*-values from a *t*-test of the pre- and posttest proficiency measures.

course because of a last-minute change at his institution; instead, students in an introductory biology course not taught by one of our participant instructors took the pre- and posttests.

Representation interview responses. Faculty members in the highest ES category (box 1) noted, first, that DQCs helped them clearly see that students could not apply concepts across biological scales and contexts. The DQC results also illuminated the students' common alternative conceptions and poor biological-reasoning abilities. Second, the descriptions of active-teaching richly show the instructors' application of such instruction across a range of courses. Third, in regard to revising courses, the availability of concrete and detailed DQC and active-learning data and information in the context of the framework on energy, matter, and scale has compelled these teachers to confront the need to reduce the amount of content in their courses and to focus on reasoning about key biological phenomena. Importantly, they have specific ideas about how to go about doing this. Finally, all of the instructors spoke in detail about various dissemination activities, and local examples are included in box 1.

Dissemination. The faculty members from four-year colleges and universities effectively used the annual ESA meeting to present talks and posters about their efforts and experiences with the project. For instance, at the 2010 meeting alone, nine DQC university or four-year-college teachers presented their work in a symposium organized by one of the instructors or in other oral presentations or posters. In contrast, the community-college faculty members presented at regional meetings such as the Association of Faculties for Advancement of Community College Teaching. Other dissemination activities included departmental talks and informal discussions with colleagues and administrators about the DQC project (box 1).

Conclusions

The DQC faculty-development program involved a range of faculty members from universities, four-year colleges, and community colleges; the participating instructors taught biology or ecology in large and small classes and devoted most of their professional time to teaching. We believe that the group fairly represents the national diversity of instructional settings and experiences of faculty members teaching these courses. Interestingly, by nearly all measures of faculty engagement and student performance, we saw few between-institution differences. Active participation in the DQC project, which included attending three annual workshops and the employment of DQC materials each year in designated courses, was generally high across the board. In written and oral surveys, the faculty members indicated substantial use of the DQCs and the associated active-learning exercises. Ten of 11 who participated in a research project saw significant pre- to posttest student gains and a large positive effect in the ES analysis; the only exception was a course not taught by one of the participating instructors. These gains, however, represent a glass half empty–glass half full situation: Although the students' ability to reason like a biologist doubled overall in response to questions concerning energy and matter for commonly taught topics in biology after they were taught these topics, Hartley and colleagues (2011) reported that only a quarter of the students achieved this advanced level of reasoning.

Box 1. What faculty in the large effect-size category said about various aspects of the Diagnostic Question Cluster (DQC) project.

Learning about students' understanding of matter and energy from the DQCs

What I liked about the DQCs… is that [they show] that when a question asks about cellular respiration [for instance] in one context they seem to get it, but not in another context.

The DQCs help me appreciate that I need to ask the same question in a different context… and ask them to compare their answers.

I honestly do not think that most of my students connect the carbon atoms studied in chemistry courses to the carbon atoms in carbon dioxide, glucose, etc.

I had expected that they would not be able to apply what they had learned at the cellular level to the ecosystem level scale, which was the case [also] many of my students believe that matter can turn into energy, and vice versa.

Most [students] could answer the question about where most of a plant's mass comes from when [the wording] was the same as we discussed in class, but when presented a different twist on the question they did not exhibit a deep understanding.

Using the DQCs and active learning exercises

I presented the [learning] activities as a worksheet that the students worked on individually and then discussed in pairs of groups of 3–4 in class.

I had them describe [their answers to a DQC question] in their own words in class.

To examine why students' extended answers were poor I interviewed seven students pre- and postinstruction to see [whether] they were simply not expressing [their scientific thinking well]… I found out that their incorrect answers made perfect sense to them, and so the problem was not their inability to explain.

I gave them [the exercise] for homework and then we talked about it the next class.

Before I started discussing the carbon cycle, I put up a carbon-flow diagram and asked them to name all of the sources of carbon to the atmosphere. I then displayed the two cow graphs ("Ecosystem Carbon Flow") and had them map out the flow of carbon in class with a partner for 10 minutes. They then answered clicker questions and then discussed it as a class.

Changing courses as a result of the project

The pretest really help[ed] me create specific activities for my students… the DQCs helped me fine tune the topics I should focus on.

I will now be able to continue to think of ways to help students confront their understanding of matter and energy relationships.

There were still many students that need to understand that matter does not turn into energy. This was very hard to help them learn and I need to figure out new ways to help them understand the relationship between matter and energy at an ecosystem and metabolic level. The instruction was there, but the learning didn't happen.

I will devote more time to the topics of energy flow and matter cycling.

How I explain concepts… I definitely need to work on clarifying some concepts. For example, I didn't [adequately] address the "used up" problem [whether energy can be used up], since there was little improvement here.

I would like to use more active-learning techniques in class. I cannot assign take-home homework with 150 students, but [the] "Think–Pair–Share" [exercise] works well for me in large lectures. Next semester, I'll incorporate a few more exercises into the discussion on carbon and energy, and I'll spend an additional class period on this material (I spent three hours this semester).

I need to reduce the number of things being tested… I did not focus enough, crammed course content.

Dissemination and rewards

I showed the data to other faculty in my department and their eyes got huge… [my chair] said that this was the first time we had evidence for student problems and… it's powerful to have evidence… for curricular reform.

I am working with another professor here, [and] we are going to coordinate to see what DQCs and active teaching we will be using.

I have been talking [to faculty members] teaching introductory biology and will be working with two of them on the DQCs.

I won [my] University's Junior Faculty Innovative Teaching Award this past academic year, partially for incorporating DQCs and activities in some of my classes… [the workshops] definitely helped to shape the proposal I submitted for the award.

The DQC program was developed in the context of examining key elements of faculty-development projects that would lead to a transformation of biology teaching and learning. In other words, what does it take for such changes to occur? Clearly this is a multidimensional issue that involves institutions and departments, among other factors, in addition to the faculty members themselves. The critical features that we have described are illustrated in figure 1. Acknowledging our focus on faculty development, we also discuss our findings in regard to four dimensions:

1. The use of research-based diagnostic conceptual inventories. A major value of the DQCs and other research-based conceptual inventories is that their formative use provides reliable, ongoing feedback about students' understanding of key concepts—critical information about the degree to which a lecture, module, or approach is working. Used as summative assessment at the end of a course, these inventories help the faculty members gauge their students' progress and ability to apply concepts more broadly. In addition, the responses are classified by a process that is useful as the faculty members improve their courses. The DQC-program participants indicated that such trusted information gave them confidence because they were more informed about aspects of the concepts that students struggle with and why. Since transformation of a course requires considerable effort in the face of time constraints, in addition to push back from colleagues and students (Handelsman et al. 2004), self-confidence is extremely important for the faculty members. In addition, pre- and posttest data from research-based instruments can be very useful in discussions with colleagues or administrators about the need for—and ways to assess—biology curricular reform.

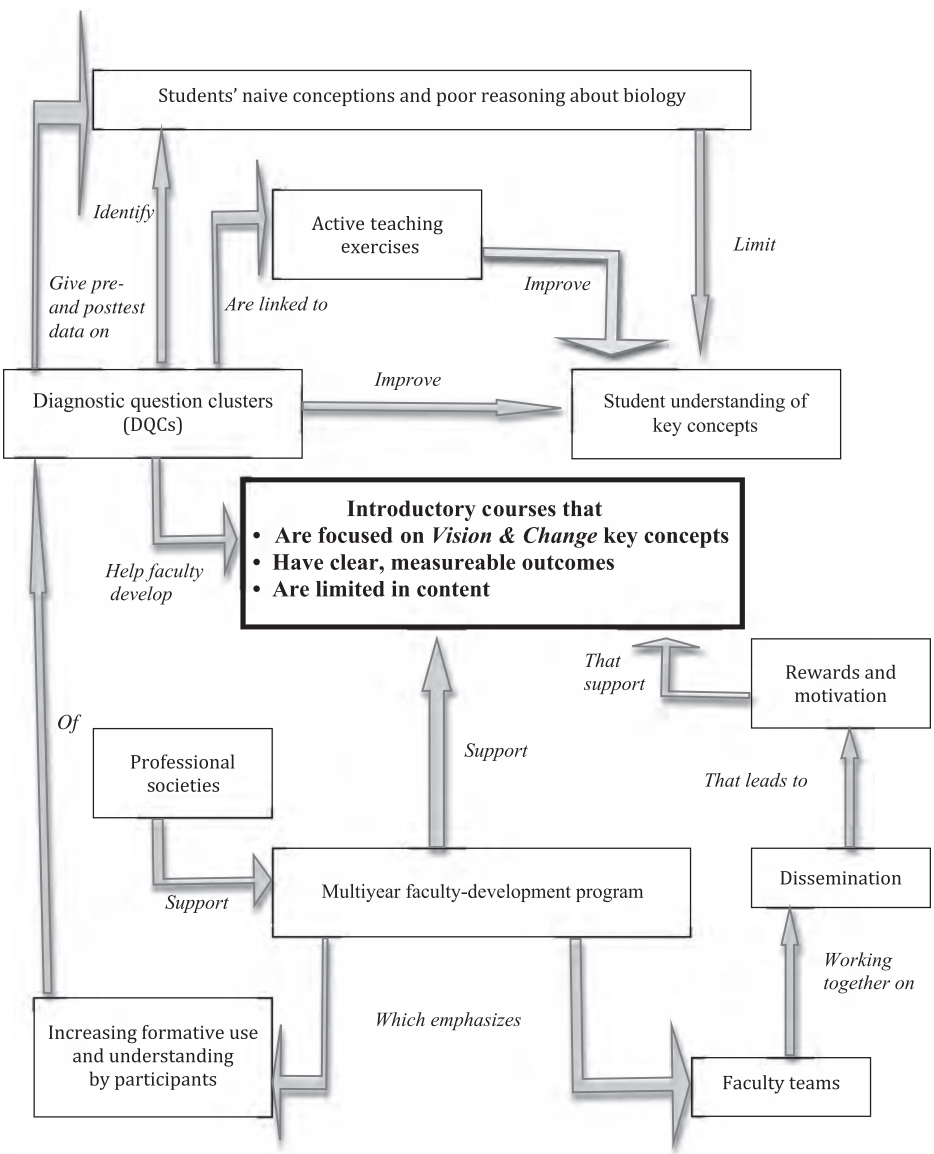

Figure 1. Key elements for faculty development in the Diagnostic Question Cluster (DQC) program that lead to a transformation of introductory biology and ecology teaching.

Faculty interviews, surveys, and discussions in workshops indicate that data from the faculty members' own classes is a critical motivator for change. Instructors may find data from other courses persuasive, but we suggest that for many, this alone is insufficient evidence that *their* students do not understand concepts and therefore that *their* teaching is somehow lacking. It is noteworthy that most of the faculty members' original approach was to wait until the end of the course to examine both the pre- and posttest data; this was partly because fitting the testing into a course was challenging at first and partly because some aimed for detachment in an unbiased, "scientific" study, instead of using the questions for the intended diagnostic purpose. Discussions in the second workshop helped the faculty members to refocus on the diagnostic purpose of the questions, so that they would better appreciate the value of using DQCs to inform their teaching and would then select active-teaching exercises to address specific problems. This stepwise development of professors' understanding over several years of how to use formative data is an important observation for faculty-development program design and is not surprising, given most faculty members' lack of experience with formative assessment and using classroom data (Yorke 2003). The time and effort required for such an evolution may be one reason that the genuine transformation of biology teaching that has been recommended in numerous reports (HHMI 1998, NRC 2003, 2009, AAAS 2011) has been limited.

Concept inventories have been used in physics, chemistry, and geology (e.g., Hestenes et al. 1992, Landis et al. 2001, Libarkin and Anderson 2006), among other sciences, in course reform. Although the development of conceptual inventories in biology is certainly growing (Haslam and Treagust 1987, Anderson DL et al. 2002, Wilson et al. 2006, Bowling et al. 2008, Garvin-Doxas and Klymkowsky 2008, Smith MK et al. 2008, Marbach-Ad et al. 2009a, Shi et al. 2010, Hartley et al. 2011), their adaptation by biology instructors is another matter. One major challenge is that biologists are a disparate group; they may be geneticists, physiologists, ecologists, and so on, which leads to disagreement about the big ideas that should be the focus of an introductory course (D'Avanzo 2007, Michael et al. 2008). We hope this will change with the publication of *Vision and Change* (AAAS 2011), which identifies five central biological concepts around which curricula and courses may be organized. Ready availability of reliable conceptual inventories is another issue, since faculty members must be able to easily obtain and compare them. In this regard, the Concept Inventory HUB (ciHUB; *http://ciHUB.org*), which is in development, is promising. This National Science Foundation–supported program will help science teachers acquire and use scientific concept inventories in order to assess student learning and to improve their pedagogical practice.

DQC faculty members have contributed to improving the DQCs—for instance, with wording and overall utility—and we view this as an ongoing process. Although some biology concept inventories were developed with considerable faculty input (e.g., Marbach-Ad et al. 2009a), many were not. We believe that faculty involvement as questions are developed, including in the validation process, will likely result in inventories appropriate for larger student audiences, among other benefits (D'Avanzo 2007).

2. Active teaching. The availability of student-active exercises was important in this project, because it is challenging and very time consuming to develop good activities that help students think more deeply about identified concepts. In addition to focusing on particular naive conceptions and difficulties (students believe that plants do not respire, or they cannot apply knowledge of cellular metabolism to organisms), the exercises were designed to help students reason about biological phenomena instead of repeating what they may have memorized for a test. The conceptual framework—the laws of conservation of energy and matter across the scales of biological organization—provided a structure for this reasoning. In year three in particular, the faculty members explained how they were deliberately using the framework to help their students become more aware of what principled reasoning is—for example, with phrases such as "tracing matter means that you…." A final critical point is that by years two and three, the faculty members were adapting exercises offered on the Web site for their courses and also developing approaches of their own.

3. Communities of practice and self-reflection. The value of learning communities is evident in numerous faculty-development programs. For example, to promote use of active teaching in a biology department, Sirum and colleagues (2009) met regularly with teams, who discussed their ongoing experiences. In developing the HPI, the faculty members validated questions by interviewing students, for instance, and discussing the students' explanations (Marbach-Ad et al. 2009a). Learning communities can be effective, because they give teachers the opportunity to step back and reflect together on their efforts to transform their teaching. Recent assessments of successful science, technology, engineering, and mathematics faculty-development programs have indicated that real change requires very active engagement of the faculty members themselves in the change process, with emphasis on self-awareness leading to action and reflective practice. For example, Henderson and colleagues (2010) described strong programs in which teachers collected classroom data that they reflected on with team members; their evidence and experience is then shared in a larger context (a poster at a meeting or a departmental talk), which increases its value to the participants. We organized teams who worked together at the first workshop and felt comfortable asking one another for advice and help during the academic year. However, the most dynamic groups consisted of faculty members who identified others that they wished to work with on posters and talks. Since few had any background in education research, trusted colleagues were essential here. These faculty members help us appreciate the necessity of flexibility and opportunity as they, like research scientists, find colleagues with whom they wish to establish close working relationships. Their experience aligns well with Rogan and Anderson's (2011) emphasis on collaborative groups working on joint projects that faculty members truly care about.

4. Motivations and rewards. For any faculty-development program to succeed, the faculty members must clearly see the value of participation—both personally and professionally—and this motivation must be strong enough to counteract strong opposing forces, such as logistics (e.g., time) and personal challenges (change is hard; Anderson TR and Rogan 2011). A principal motivation built into the DQC program—presentation at meetings—has been effective when the faculty members attend conferences that they care about with colleagues who are interested in their work. For ecologists at universities and four-year colleges, the ESA meeting has served this purpose well. For some, these presentations are considered in reappointment and promotion decisions—clearly important incentives for the faculty members. In contrast, community-college faculty members appear to value informal, regional meetings more than national ones. Understanding the importance of a variety of dissemination venues is a valuable outcome of our work. Interestingly, this was the only observable difference between community college and four-year college and university faculty.

Program expansion and sustainability. One of the most rewarding outcomes of the project has been the emergence of a new generation of biology and ecology education leaders; in this program, nine assistant and associate professors have presented posters and talks, led symposia, and given workshops at one ESA meeting alone. In addition, these faculty members gave DQC presentations at their institutions, they worked individually with colleagues on DQCs, and some even recruited other faculty members for future projects. From the beginning, these instructors were engaged as partners in a community of educators striving to improve the teaching and learning of biology and ecology in colleges and universities nationwide. Clearly, intensive professional-development programs like this one cannot be sustained without an expanding community of leaders—faculty participants who then lead programs of their own using the same or similar tools, resources, pedagogy, and approaches to faculty development. This finding supports the notion that professional societies, such as the ESA, can be very effective venues to support such endeavors. Numerous biology societies are already actively engaged in pedagogical reform and thus poised to take on the *Vision and Change* (AAAS 2011) challenges. We propose that research-based biology conceptual inventories focused on the concepts and topics that faculty members typically teach; on the explicit formative use of diagnostic data in course reform; on the communities of practice, with faculty members working together over several years; and on scholarly dissemination by the instructors can be very effective in these faculty-development programs.

Acknowledgments

We thank the instructors who administered the diagnostic question clusters to their students and who participated in helpful discussions related to their students' understanding of matter and energy: Peter Armbruster, Amy Arnett, Barbara Abraham, Patrick Crumrine, Alan Dickman, Heather Griscom, Casey Huckins, April Maskiewicz, Beth Olson, Chris Picone, Nicole Welch. We also thank our collaborators on the project, who include Jing Chen, Jennifer Doherty, Alan Griffith, John Merrill, Johanthan Schramm, Nancy Stamp, and Kathy Williams. April Maskiewicz helped with revisions of the manuscript. This project was supported by National Science Foundation Division of Undergraduate Education awards 0919992, 0920186, and 0127388.

References cited

[AAAS] American Association for the Advancement of Science. 2011. Vision and Change in Introductory Biology Education: A Call to Action. AAAS.

Anderson DL, Fisher KM, Norman GJ. 2002. Development and evaluation of the conceptual inventory of natural selection. Journal of Research in Science Teaching 39: 952–978.

Anderson TR, Rogan JM. 2011. Bridging the educational research-teaching practice gap: Curriculum development, part 1: Components of the curriculum and influences on the process of curriculum design. Biochemistry and Molecular Biology Education 39: 68–76.

Bowling BV, Acra EE, Wang L, Myers MF, Dean GE, Markle GC, Moskalik CL, Huethe CA. 2008. Development and evaluation of a genetics literacy assessment instrument for undergraduates. Genetics 178: 15–22.

Chi MTH, Feltovich PJ, Glaser R. 1981. Categorization and representation of physics problems by experts and novices. Cognitive Science 5: 121–152.

Cohen J. 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Erlbaum.

Cohen L, Manion L, Morrison K. 2000. Research Methods in Education, 5th ed. Routledge.

D'Avanzo C. 2007. Changing Teaching Practice: Much More than a Diagnostic Test. Paper presented at the first Conceptual Assessment in Biology meeting; 3–4 March 2007, Boulder, Colorado. (1 February 2012; *bioliteracy.colorado.edu/CABS.html*)

———. 2008. Biology concept inventories: Overview, status, and next steps. BioScience 58: 1079–1085.

Ebert-May D, Derting TL, Hodder J, Momsen JL, Long TM, Jardeleza SE. 2011. What we say is not what we do: Effective evaluation of faculty professional development programs. BioScience 61: 550–558.

Garvin-Doxas K, Klymkowsky MW. 2008. Understanding randomness and its impact on student learning: Lessons learned from building the Biology Concept Inventory (BCI). CBE Life Science Education 7: 227–233.

Garvin-Doxas K, Klymkowsky M[W], Elrod S. 2007. Building, using, and maximizing the impact of concept inventories in the biological sciences: Report on a National Science Foundation sponsored conference on the construction of concept inventories in the biological sciences. CBE Life Science Education 6: 277–282.

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

Hartley LM, Wilke BJ, Schramm JW, D'Avanzo C, Anderson CW. 2011. College students' understanding of the carbon cycle: Contrasting principle-based and informal reasoning. BioScience 61: 65–75.

Haslam F, Treagust DF. 1987. Diagnosing secondary students' misconceptions of photosynthesis and respiration in plants using a two-tier multiple choice instrument. Journal of Biology Education 21: 203–211.

Henderson C, Finkelstein N, Beach A. 2010. Beyond dissemination in college *science* teaching: An introduction to four core change strategies. Journal of College Science Teaching 39: 18–25.

Hestenes D, Wells M, Swackhamer G. 1992. Force concept inventory. The Physics Teacher 30: 141–158.

[HHMI] Howard Hughes Medical Institute. 1998. Beyond Bio 101: The Transformation of Undergraduate Biology Education. HHMI. (1 February 2012; *www.hhmi.org/BeyondBio101*)

Knight JK, Wood WB. 2005. Teaching more by lecturing less. Cell Biology Education 4: 298–310.

Landis CR, Ellis AB, Lisensky GC, Lorenz JK, Meeker K, Wamser CC. 2001. Chemistry ConcepTests: A Pathway to Interactive Classrooms. Prentice Hall.

Libarkin JC, Anderson SW. 2006. The geoscience concept inventory: Application of Rasch analysis to concept inventory development in higher education. Pages 45–73 in Liu X, Boone W, eds. Applications of Rasch Measurement in Science Education. JAM Press.

Marbach-Ad G, et al. 2009a. Assessing student understanding of host pathogens using a concept inventory. Journal of Microbiology Education 10: 43–50.

Marbach-Ad, et al. 2009b. A model for using a concept inventory as a tool for students' assessment and faculty professional development. CBE Life Sciences Education 9: 408–416.

Masters GN. 1982. A Rasch model for partial credit scoring. Psychometrika 47: 149–174.

Michael JA, Richardson D, Rovick A, Modell H, Bruce D, Horowitz B, Hudson M, Silverthorn D, Whitescarver S, Williams S. 1999. Understanding students' misconceptions about respiratory physiology. Advances in Physiological Education 22: S127–S135.

Michael JA, McFarland J, Wright A. 2008. The second Conceptual Assessment in the Biological Sciences workshop. Advances in Physiological Education 32: 248–251.

Mohan L, Chen J, Anderson CW. 2009. Developing a multi-year learning progression for carbon cycling in socio-ecological systems. Journal of Research in Science Teaching 46: 675–698.

Momsen JL, Long TW, Wyse SA, Ebert-May D. 2010. Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. CBE Life Science Education 9: 435–440.

[NRC] National Research Council. 2003. BIO 2010: Transforming Undergraduate Education for Future Research Biologists. National Academies Press.

———. 2009. A New Biology for the 21st Century. National Academies Press.

[PKAL] Project Kaleidoscope. 1991. What Works: Building Natural Science Communities. PKAL. (1 February 2012; *www.aacu.org/pkal/publications/index.cfm*)

Rogan JM, Anderson TR. 2011. Bridging the educational research–teaching practice gap: Curriculum development, part 2: Becoming an agent of change. Biochemistry and Molecular Biology Education 39: 233–241.

Shi J, Wood WB, Martin JM, Guild NA, Vicens Q, Knight JK. 2010. A diagnostic assessment for introductory molecular and cell biology. CEB Life Science Education 9: 453–461.

Sirum KL, Madigan D, Klionsky DJ. 2009. Enabling a culture of change: A life science faculty learning community promotes scientific teaching. Journal of College Science Teaching 38: 24–30.

Smith JI, Tanner K. 2010. The problem of revealing how students think: Concept inventories and beyond. CBE Life Science Education 9: 1–5.

Smith MK, Wood WB, Knight JK. 2008. The Genetics Concepts Assessment: A new concept inventory for gauging student understanding of genetics. CBE Life Science Education 7: 422–430.

Tanner K, Allen D. 2005. Approaches to biology teaching and learning: Understanding the wrong answers—teaching toward conceptual change. CBE Life Science Education 4: 112–117.

Vaske JJ, Gliner JA, Morgan GA. 2002. Communicating judgments about practical significance: Effect size, confidence intervals, and odd ratios. Human Dimensions of Wildlife 7: 287–300.

Wilson CD, Anderson CW, Heidemann M, Merrill JE, Merritt BW, Richmond G, Sibley DF, Parker JM. 2006. Assessing students' ability to trace matter in dynamic systems in cell biology. CBE Life Science Education 5: 323–331.

Yorke M. 2003. Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. Higher Education 45: 477–501.

Charlene D'Avanzo (cdavanzo@hampshire.edu) is a professor in the School of Natural Sciences at Hampshire College, in Amherst, Massachusetts. Charles W. Anderson is a professor in the Department of Teacher Education at Michigan State University, in East Lansing. Laurel M. Hartley is an assistant professor in the Department of Integrative Biology at the University of Colorado Denver. Nancy Pelaez is an associate professor of biological science at Purdue University, in West Lafayette, Indiana.