## HIGHLIGHTS OF 2013

### Table of Contents

**Editorial**

**A Year of Firsts**
Erin L. Dolan ................................................................. 1–2

**FEATURES**

**Approaches to Biology Teaching and Learning**

**Structure Matters: Twenty-One Teaching Strategies to Promote Student Engagement and Cultivate Classroom Equity**
Kimberly D. Tanner ...................................................... 3–12

**Barriers to Faculty Pedagogical Change: Lack of Training, Time, Incentives, and . . . Tensions with Professional Identity?**
Sara E. Brownell and Kimberly D. Tanner ................. 13–20

**WWW. Life Sciences Education**

**Physics and Biology Collaborate to Color the World**
Dennis W. C. Liu ............................................................. 21–26

**From the National Research Council**

**Biology Education Research: Lessons and Future Directions**
Susan R. Singer, Natalie R. Nielsen, and Heidi A. Schweingruber .................................................. 27–30

**Book Review**

**Genesis of What Is Life?: A Paradigm Shift in Genetics History**
John R. Jungck ............................................................... 31–32

**RESEARCH METHODS**

**The Other Half of the Story: Effect Size Analysis in Quantitative Research**
Jessica Middlemis Maher, Jonathan C. Markey, and Diane Ebert-May .................................................. 33–39

**ESSAY**

**The Teaching Demonstration: What Faculty Expect and How to Prepare for This Aspect of the Job Interview**
Michelle K. Smith, Mary Pat Wenderoth, and Mary Tyler ................................................................. 40–46

*Table of Contents Continued on Back Cover*
NEW FOURTH EDITION
ESSENTIAL CELL BIOLOGY

Bruce Alberts, University of California, San Francisco, USA
Dennis Bray, University of Cambridge, UK
Karen Hopkin, Science Journalist and Biochemist, USA
Alexander Johnson, University of California, San Francisco, USA
Julian Lewis, Emeritus, Cancer Research UK, UK
Martin Raff, University College London, UK
Keith Roberts, Emeritus, John Innes Centre, UK
Peter Walter, University of California, San Francisco, USA

October 2013 • 864 pages • 860 illustrations
Hardback • 978-0-8153-4454-4 • $151

Essential Cell Biology provides a readily accessible introduction to the central concepts of cell biology, and its lively, clear writing and exceptional illustrations make it the ideal textbook for a first course in both cell and molecular biology. The text and figures are easy-to-follow, accurate, clear, and engaging for the introductory student. Molecular detail has been kept to a minimum in order to provide the reader with a cohesive conceptual framework for the basic science that underlies our current understanding of all of biology, including the biomedical sciences.

The Fourth Edition has been thoroughly revised, covers the latest developments in this fast-moving field, yet retains the academic level and length of the previous edition. The book is accompanied by a rich package of online student and instructor resources, including over 130 narrated movies, an expanded and updated Question Bank, and new enhanced assessments for students.

TO ORDER IN THE US, CANADA, AND LATIN AMERICA, CONTACT:

Taylor & Francis
7625 Empire Drive
Florence, KY 41042
Call toll-free: 1-800-634-7064
Call International: 561-361-6000 x 6418
Fax toll-free: 1-800-248-4724
Fax International: 561-361-6075
E-mail: orders@taylorandfrancis.com

For ordering information from other regions, or to request an examination copy, e-mail: science@garland.com

NEW IN THE FOURTH EDITION

• Comprehensive scientific update of the entire text and figures in the book.
• New material on regulatory RNAs, epigenetics, induced pluripotent stem cells, cell suicide and reprogramming, the human genome, and surprising inheritance from Neanderthals.
• Based on extensive student feedback, improved discussion and new figures of photosynthesis, metabolism, and DNA repair.
• Updated coverage of many exciting, new experimental techniques including RNAi, optogenetics, the applications of new DNA sequencing technologies, and the use of mutant organisms to probe the defects underlying human disease.
• Medically relevant topics include new sections on culturing human cells for clinical work, autism and iPSC cells, peroxisomes and disease, cancer changes in key regulatory pathways, as well as expanded coverage of amyloid proteins and their role in Alzheimer’s and Huntington’s diseases, the use of genome sequencing to identify the rare mutations that predispose us to disease, and much more.
• New and enhanced student self-assessments available online.

Please visit us at the annual meeting of the American Society for Cell Biology at Booth 217 to browse the new Fourth Edition of Essential Cell Biology, and to help us celebrate the 30th anniversary of the publication of Molecular Biology of the Cell.
NEW! iBioEducation

iBioEducation offers something for every student with short clips, lectures, classroom materials, as well as a microscopy course.

Microscopy Course:
Over 80 lectures, lab demos, and tips for using a light microscope by microscopy experts.

iBioSeminars

Over 85 biology seminars given by the world’s leading researchers. Each lecture includes both a broad overview of the field and research intensive talks.

iBioMagazine

Short talks that go behind the science, featuring topics such as science policy, careers, education, advice, science and society, how major scientific discoveries were made, and a series on “How I Became a Scientist.”

All this and more is available on our new website: www.iBiology.org
For educators at all levels and across all life science disciplines.

Online and completely FREE. No subscription required.

LSE emphasizes teaching innovations and evidence of their effectiveness. It publishes original research articles, essays, and features that help you apply education research to your own teaching.

Published by ascb
the american society for cell biology

Editorial Partner GSA
Genetics Society of America

Funded in part by the Howard Hughes Medical Institute
Postdocs/Grad Students

Do you want to

Organize a One-Day Local Meeting?

ASCB Financial Support Available

Take advantage of this career advancement opportunity from ASCB. Such meetings will typically involve two or more local research institutions or colleges (within or outside of the USA). Topics can range from basic science to career development, as long as there is clear relevance to the broadly defined field of cell biology.

For more information go to www.ascb.org and click on “Meetings”, then “Local Meetings” or email ssalzer@ascb.org.

Next Deadline for Applications:
February 1, 2013
A Year of Firsts

Erin L. Dolan

University of Georgia, Athens, GA 30602

This has been a year of firsts for CBE—Life Sciences Education (LSE). LSE was the first journal to publish a collection of education research and practice articles integrating physics and biology education. Articles that comprise the Integrating Physics and Biology issue take important steps toward responding to the call for teaching and learning at the intersection of biology and physics, and raise interesting and significant questions for future research and practice in physics and biology education research. LSE published its first research methods essay, the first in a series intended to translate social science methods in ways that are compelling to and approachable for trained biologists. LSE is also publishing articles from an increasingly international group of authors on a broader range of issues in biology education—from how to effectively implement professional development on teaching, to how to measure knowledge about fundamental biological ideas. LSE’s growth has spurred us to initiate continuous publication in 2014, such that articles accepted for publication are available as quickly as possible rather than held for four predefined dates.

Several firsts have occurred behind the scenes as well. Our publisher, the American Society for Cell Biology (ASCB), has established a fund to ensure the long-term sustainability of the journal without resorting to subscription fees or page charges. ASCB has seeded the fund with an initial commitment of $100,000. This novel commitment of funds demonstrates that ASCB is serious about sustaining the journal and disseminating current knowledge about biological ideas. LSE’s growth has spurred us to initiate continuous publication in 2014, such that articles accepted for publication are available as quickly as possible rather than held for four predefined dates.

Where will the biology education community be after all of these firsts? Teachers will be better at articulating learning objectives that are meaningful and measurable. Students will benefit from assessments that align with clear learning objectives. Courses will engage students in learning activities that align with learning objectives, reveal where students are in their learning, and prepare students to be successful on summative assessments. Homework and projects will provide our students with opportunities to verbalize their thinking, reflect on their knowledge, engage in meaningful interactions, and receive informative feedback. Effective teaching begins by assessing students—what do they know and how are they thinking? Effective instruction provides students with opportunities to practice the kind of thinking we expect from them and that they will need to succeed in STEM education and careers. However, not all faculty embrace evidence-based teaching, because change is difficult, change takes time, and change requires a shift in beliefs and culture (Gess-Newsome...
I encourage you to accomplish some firsts of your own to mentor others in the use of evidence-based instruction. Invite a colleague to your class to watch you teach. Ask him or her to focus on what the students are doing and how your teaching prompts students to actively engage in learning. Share an LSE article with someone and offer to discuss the article over coffee. Volunteer to teach a lesson in a colleague’s class to demonstrate active learning and what students can achieve when given the opportunity. Gather a subset of your colleagues to take turns observing one another’s classes and giving feedback to help improve student learning. Involve an undergraduate, graduate, or postdoctoral researcher in collecting assessment data in one of your classes and developing strategies to reduce student confusion or lack of mastery of core concepts. Share the PULSE rubrics with your department head and start a conversation about where your department is and how to make progress in adopting evidence-based teaching practices. Implement a new strategy or intervention that has been shown to be effective. Put your students’ learning first and join the global community as we celebrate the ever-expanding examples of educational firsts.

REFERENCES


Feature
Approaches to Biology Teaching and Learning

Structure Matters: Twenty-One Teaching Strategies to Promote Student Engagement and Cultivate Classroom Equity
Kimberly D. Tanner
Department of Biology, San Francisco State University, San Francisco, CA 94132

INTRODUCTION
As a biology education community, we focus a great deal of time and energy on issues of “what” students should be learning in the modern age of biology and then probing the extent to which students are learning these things. Additionally, there has been increased focus over time on the “how” of teaching, with attention to questioning the efficacy of traditional lecture methods and exploring new teaching techniques to support students in more effectively learning the “what” of biology. However, the aspect of classroom teaching that seems to be consistently underappreciated is the nature of “whom” we are teaching. Undergraduate students often appear to be treated as interchangeable entities without acknowledgment of the central role of the individual students, their learning histories, and their personal characteristics in the student-centered nature of “how” we aspire to teach. Most innovative approaches to biology teaching that are at the core of national policy documents and resources are rooted in a constructivist framework (e.g., Posner et al., 1982; Handelsman et al., 2004; Labov et al., 2010; American Association for the Advancement of Science [AAAS], 2011; College Board, 2013). In constructivism, teachers can structure classroom environments with the intention of maximizing student learning, but learning is the work of students (Posner et al., 1982; Bransford et al., 2000). As such, each student’s prior experience and attitude and motivation toward the material being learned, confidence in his or her ability to learn, and relative participation in the learning environment are all thought to be key variables in promoting learning of new ideas, biological or not. Finally, bringing together individual students in classrooms produces group interactions that can either support or impede learning for different individuals.

Designing learning environments that attend to individual students and their interactions with one another may seem an impossible task in a course of 20 students, much less a course of more than 700. However, there are a host of simple teaching strategies rooted in research on teaching and learning that can support biology instructors in paying attention to whom they are trying to help learn. These teaching strategies are sometimes referred to as “equitable teaching strategies,” whereby striving for “classroom equity” is about teaching all the students in your classroom, not just those who are already engaged, already participating, and perhaps already know the biology being taught. Equity, then, is about striving to structure biology classroom environments that maximize fairness, wherein all students have opportunities to verbally participate, all students can see their personal connections to biology, all students have the time to think, all students can pose ideas and construct their knowledge of biology, and all students are explicitly welcomed into the intellectual discussion of biology. Without attention to the structure of classroom interactions, what can often ensue is a wonderfully designed biology lesson that can be accessed by only a small subset of students in a classroom.

So what specific teaching strategies might we instructors, as architects of the learning environment in our classrooms, use to structure the classroom learning environment? Below are 21 simple teaching strategies that biology instructors can use to promote student engagement and cultivate classroom equity. To provide a framework for how these teaching strategies might be most useful to instructors, I have organized them into five sections, representing overarching goals instructors may have for their classrooms, including:

- Giving students opportunities to think and talk about biology
- Encouraging, demanding, and actively managing the participation of all students
- Building an inclusive and fair classroom community for all students
- Monitoring behavior to cultivate divergent biological thinking
- Teaching all of the students in your biology classroom

DOI: 10.1187/cbe.13-06-0115
Address correspondence to: Kimberly Tanner (kdtanner@sfsu.edu).
© 2013 K. D. Tanner. CBE—Life Sciences Education © 2013 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (http://creativecommons.org/licenses/by-nc-sa/3.0/).
“ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.
For each of these goals, there is a brief consideration of why the goal is important for student learning, which is followed by descriptions of several simple strategies for structuring instructor–student and student–student interactions to strive for this goal. No doubt, there are likely dozens of additional strategies that could be added to this list. In addition, many of the strategies affiliated with one equitable teaching goal are also easily used in the service of one or more of the other goals. The intention of presenting these 21 strategies in this framework is solely to provide all biology instructors access to immediate and tractable teaching strategies for promoting access and equity for all students in their biology classrooms.

These equitable teaching strategies can be read and explored in any order. Readers are encouraged to use Table 1 to self-assess which of these strategies they may already use, which they are most interested in reading more about, and which they may want to try in their own classrooms. Self-assessment responses to Table 1 can guide which of the sections below you may be most interested in reading first.

### GIVING STUDENTS OPPORTUNITIES TO THINK AND TALK ABOUT BIOLOGY

Human learning is a biological phenomenon of the brain. Synapses need time to fire, and relevant circuits in the brain need time to be recruited. Yet the structure of class time with students does not usually attend to giving students time to think and talk about biology. As experts with thousands of hours of thinking about biology, we as biologists no doubt think quite quickly about the topics we are attempting to teach students. And we as instructors can be misled that all students have had ample time to think by those few students in our courses who have more background in the concepts under discussion and raise their hands to share almost immediately. However, those students in our courses who are more biologically naïve may need more time to think and talk about the biological concepts under discussion. Below are four simple teaching strategies grounded in research to structure classroom time for students to think and talk about biology.

#### 1. Wait Time

Perhaps the simplest teaching strategy to increase time for student thinking and to expand the number of students participating verbally in a biology classroom is to lengthen one’s “wait time” after posing a question to your class (Rowe, 1969; Tobin, 1987). Mary Budd Rowe’s groundbreaking papers introducing the concept of wait time have influenced educational practice since their publication more than 40 years ago (Rowe, 1969, 1974, 1978; Tannen and Allen, 2002). Rowe and colleagues documented in the precollege setting that instructors on average waited only ~1.5 s after asking a question before taking a student response, answering the question themselves, or posing a follow-up question. With the seemingly modest extension of the “wait time” after a question to ~3–5 s, Rowe and colleagues showed dramatic effects: substantially more students willing to volunteer answers, fewer students unwilling to share when called on, and increases in the length and complexity of the responses that students gave in response to the question (Rowe, 1974; Allen and Tanner, 2002). Thinking biologically about increasing wait time to promote student engagement and participation, it seems likely that this increase in time allows critical neural processing time for students, and perhaps also allows more introverted students time to rally the courage to volunteer an answer. Practically, extending wait time can be very challenging for instructors. Actively mentally counting the following—“one thousand one . . . one thousand two . . . one thousand three . . . one thousand four . . . one thousand five”—before acknowledging potential student respondents is one simple way to track the amount of time that has transpired after asking a question.

#### 2. Allow Students Time to Write

Practicing wait time may still not give enough time for some students to gather a thought and or screw up the confidence to share that thought. Many students may need more scaffolding—more instruction and guidance—about how to use the time they have been given to think. One simple way to scaffold wait time is to explicitly require students to write out one idea, two ideas, three ideas that would capture their initial thoughts on how to answer the question posed. This act of writing itself may even lead students to discover points of confusion or key insights. In addition, if collected, this writing can hold students accountable in thinking and recording their ideas. To set the stage for doing these simple quick writes or minute papers throughout the semester, instructors can require on the syllabus that students purchase a packet of index cards (usually no more than a $1 cost) and bring a few cards to each class session for the purpose of these writing opportunities. Instructors need not collect all of these writings, though it may be quite informative to do so, and certainly instructors need not grade any (much less every) card that students produce. If these quick writes are graded, it can be only for participation points or more elaborately to provide conceptual feedback (Schinske, 2011). Giving students time to write is one way that instructors can structure the learning environment to maximize the number of students who have access (in this case enough time) to participate in thinking about biology.

#### 3. Think–Pair–Share

The oft written about think–pair–share strategy is perhaps the simplest way for instructors coming from a traditional lecture approach to give all students in a classroom opportunities to think about and talk about biology (Lyman, 1981; Chi et al., 1994; Allen and Tanner, 2002; Smith et al., 2009; Tanner, 2009). The mechanics of a think–pair–share generally involve giving all students a minute or so to think (or usually write) about their ideas on a biological question. Then, students are charged to turn and talk with a neighboring student, compare ideas, and identify points of agreement and misalignment. These pair discussions may or may not be followed by a whole-group conversation in which individual students are asked to share the results of their pair discussion aloud with the whole class. Importantly, the instructor’s role in facilitating a think–pair–share activity is to be explicit that students need not agree and also to convey that practicing talking about biology is an essential part of learning about biology. Integrating one or more think–pair–share opportunities...
Table 1. Self-assessment of equitable teaching strategies\textsuperscript{a}

<table>
<thead>
<tr>
<th>Giving students opportunities to think and talk about biology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Wait time</td>
</tr>
<tr>
<td>2. Allow students time to write</td>
</tr>
<tr>
<td>3. Think–pair–share</td>
</tr>
<tr>
<td>4. Do not try to do too much</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Encouraging, demanding, and actively managing the participation of all students</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Hand raising</td>
</tr>
<tr>
<td>6. Multiple hands, multiple voices</td>
</tr>
<tr>
<td>7. Random calling using popsicle sticks/index cards</td>
</tr>
<tr>
<td>8. Assign reporters for small groups</td>
</tr>
<tr>
<td>9. Whip (around)</td>
</tr>
<tr>
<td>10. Monitor student participation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Building an inclusive and fair biology classroom community for all students</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Learn or have access to students’ names</td>
</tr>
<tr>
<td>12. Integrate culturally diverse and relevant examples</td>
</tr>
<tr>
<td>13. Work in stations or small groups</td>
</tr>
<tr>
<td>14. Use varied active-learning strategies</td>
</tr>
<tr>
<td>15. Be explicit about promoting access and equity for all students</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Monitoring (your own and students’) behavior to cultivate divergent biological thinking</th>
</tr>
</thead>
<tbody>
<tr>
<td>16. Ask open-ended questions</td>
</tr>
<tr>
<td>17. Do not judge responses</td>
</tr>
<tr>
<td>18. Use praise with caution</td>
</tr>
<tr>
<td>19. Establish classroom community and norms</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Teaching all of the students in your biology classroom</th>
</tr>
</thead>
<tbody>
<tr>
<td>20. Teach them from the moment they arrive</td>
</tr>
<tr>
<td>21. Collect assessment evidence from every student, every class</td>
</tr>
</tbody>
</table>

Other equitable teaching strategies I use …

\textsuperscript{a}Spaces to the left of each strategy can be used to indicate: N = never used; O = use occasionally; R = use regularly; W = would like to try!
4. Do Not Try to Do Too Much

Finally, no instructors would likely express the sentiment: “I try to do so much in my class sessions that they go by quickly and students are unclear about what the goals for the class were.” However, evidence from a variety of research studies suggests that this may be the dominant experience for many students in undergraduate science courses (Tobias, 1990; Seymour and Hewitt, 1997). While “not doing too much” is a challenging task for most of us, one particular strategy that can reduce the amount of material considered during class time is to structure more active learning by students outside class time, in particular in the form of homework that goes beyond textbook readings. Examples include case study assignments that charge students to independently explore and find evidence about an upcoming conceptual idea before arriving in class. As experts in our biological fields, it is tempting to continually expand what we deem critical and nonnegotiable in terms of what students need to accomplish during class time. However, there are clear and present trade-offs between continually expanding our aspirations for in-class time and structuring a classroom learning environment that promotes student engagement and provides access to thinking and talking about biology for all students. One strategy for prioritizing how to spend precious class time is to decide on which biological ideas in a course are most difficult to learn, are rooted in common misconceptions, and/or represent fundamental biological principles (National Research Council, 1999; AAAS, 2011; Coley and Tanner, 2012).

ENCOURAGING, DEMANDING, AND ACTIVELY MANAGING THE PARTICIPATION OF ALL STUDENTS

If learning requires that students construct ideas for themselves, then demanding the active participation of every single student in a class is essential to learning. Currently, though, many undergraduate students in biology classrooms can navigate an entire term without speaking aloud in a course. They sit in the back of our large classrooms, and they attempt to appear to be busily writing when a question is asked in a small class. Being called upon to answer a question or share an idea can be deeply uncomfortable to many students, and we as instructors may not be doing enough to build students’ confidence to share. While few instructors would find this lack of active, verbal participation in science acceptable for emerging scientists such as graduate students or practicing scientists themselves, we somehow allow this for undergraduate students. The participation of a few students in our classrooms on a regular basis, often from the front rows, distacts us from the fact that usually the vast majority of students are not participating in the conversation of biology. To encourage, and in fact demand, the participation of all students in a biology classroom, you can use the following six strategies with little to no preparation or use of class time.

5. Hand Raising

Actively enforcing the use of hand raising and turn taking in a classroom is likely to provide greater access to more students than an open, unregulated discussion. Novice instructors, sometimes awash in silence and desperate for any student participation, can allow the classroom to become an open forum. Some would say this is much like the culture of science in settings such as lab meetings and seminars. However, the undergraduates in our courses are novices, not only to the concepts we are sharing but also to the culture of science itself. As such, providing structure through something as simple as hand raising can establish a culture that the instructor expects all students to be participating. With hand raising, the instructor can also be explicit about asking for “hands from those of us who haven’t had a chance yet to share” and strive to cultivate a classroom conversation that goes beyond a few students in the front row.

6. Multiple Hands, Multiple Voices

After asking a question, some instructors call on just a single student to answer. However, this is problematic in many ways. The same students can often end up sharing repeatedly during a class, as well as from class session to class session. In addition, if the goal is to better understand how students are thinking, having a single student share gives a very narrow and highly skewed picture of what a classroom full of students may be thinking. One simple strategy for broadening participation and increasing the breadth of ideas flowing from students to instructors is to generally ask for multiple hands and multiple voices to respond to any question posed during class time (Allen and Tanner, 2002). Instructors can set the stage for this by asserting, “I’m going to pose a question, and I’d like to see at least three hands of colleagues here who would share their ideas. I won’t hear from anyone until I’ve got those three volunteers.” Additionally, this particular use of hand raising allows instructors to selectively call on those students who may generally participate less frequently or who may have never previously shared aloud in class. Importantly, instructors really must always wait for the number of hands that they have called for to share. Hearing from fewer than the number of volunteers called for can entrain students in a classroom to know that they simply have to outwait the instructor. Finally, if the number of requested hands have not been volunteered, the instructor can charge students to talk in pairs to rehearse what they could share if called upon to do so.
7. Random Calling Using Popsicle Sticks/Index Cards

Raising hands allows for the instructor to structure and choose which students are participating verbally in a class, but what if no one is raising a hand or the same students continually raise their hands? Establishing the culture in a classroom that any student can be called on at any time is another option for promoting student engagement and participation. How this is done can be critical. If the spirit of calling on students feels like a penalty, it may do more harm than good. However, if the instructor is explicit that all students in the course have great ideas and perspectives to share, then random calling on students in courses that range in size from 10 to 700 can be a useful strategy for broadening student participation. Practically, there are a variety of ways to call randomly on students. In smaller-sized courses, having a cup with popsicle sticks, each with the name of a student on it, can make the process transparent for students, as the instructor can clearly hold up the cup, draw three names, read the names, and begin the sharing. This can minimize suspicions that the instructor is preferentially calling on certain students. For larger course class sizes, instructors can collect an index card with personal information from each student on the first day. The cards serve two purposes: 1) to enable instructors to get to know students and to assist with learning students’ names, and 2) to provide a card that can be used each class and cycled through over the semester to randomly call on different students to share (Tanner, 2011).

8. Assign Reporters for Small Groups

Promoting student engagement and classroom equity involves making opportunities for students to speak who might not naturally do so on their own. If the decision about who is to share aloud in a class discussion is left entirely to student negotiation, it is no surprise that likely the most extroverted and gregarious students will repeatedly and naturally jump at all opportunities to share. However, this sets up an inequitable classroom environment in which students who are unlikely to volunteer have no opportunities to practice sharing their scientific ideas aloud. Assigning a “reporter”—an individual who will report back on their small-group discussion—is a simple strategy to provide access to verbal participation for students who would not otherwise volunteer. The assignment of reporters need not be complex. It can be random and publicly verifiable, such as assigning that the reporter will be the person wearing the darkest shirt. In smaller classes, one can use simple tools to assign a reporter, such as colored clips on individual student name tents or colored index cards handed to students as they enter the class. It can also be nonrandom and intended to draw out a particular population. For example, assigning the group reporter to be the person with the longest hair will often, not always, result in a female being the reporter for a group. Or instructors can choose to hand out the colored clips/cards specifically to students who are less likely to share their ideas in class. Early on, it may be useful to assign based on a visible characteristic, so the instructor can verify that those students reporting are indeed those who were assigned to report. After the culture of assigned reporters is established, and everyone is following the rules, assignments can become less verifiable and prompt more personal sharing, such as the reporter is the person whose birthday is closest. Whatever the method, assigning reporters is a simple strategy for promoting classroom fairness and access to sharing ideas for more than just the most extroverted students.

9. Whip (Around)

Actively managing the participation of all students in smaller courses is sometimes well supported by the occasional use of what is termed a “whip around” or more simply just a “whip.” In using a whip, the instructor conveys that hearing an idea from every student in the classroom is an important part of the learning process. Whips can be especially useful toward the beginning of a course term as a mechanism for giving each student practice in exercising his or her voice among the entire group, which for many students is not a familiar experience. The mechanics of the whip are that the instructor poses a question to which each individual student will respond, with each response usually being ~30 s in length. On the first day of class, this could be something as simple as asking students what their favorite memory of learning biology has been. As the course progresses, the question that is the focus of the whip can become more conceptual, but always needs to be such that there are a variety of possible responses. Whips can be follow-ups to homework assignments wherein students share a way in which they have identified a personal connection to course material, a confusion they have identified, or an example of how the material under study has recently appeared in the popular press. During a whip, students who may wish to share an idea similar to a colleague who has previously shared are actively encouraged to share that same idea, but in their own words, which may be helpful to the understanding of fellow students or reveal that the ideas are not actually that similar after all. Importantly, the whip is a teaching strategy that is not feasible in large class sizes, as the premise of the strategy is that every student in the class will respond. As such, this strategy is unwieldy in class sizes greater than ~30, unless there is a subgroup structure at play in the classroom with students already functioning regularly in smaller groups. Possible ways to implement a whip in a large classroom could be to call on all students in a particular row or in a particular subgroup structure particular to the course.

10. Monitor Student Participation

Many instructors are familiar with collecting classroom evidence to monitor students’ thinking, using clicker questions, minute papers, and a variety of other assessment strategies. Less discussed is the importance of monitoring students’ participation in a classroom on a regular basis. It is not unusual to have a subset of students who are enthusiastic in their participation, sometimes to the point that the classroom dialogue becomes dominated by a few students in a room filled with 20, 40, 80, 160, or upward of 300 students. To structure the classroom dialogue in such a way as to encourage, demand, and actively manage the participation of all students, instructors can do a variety of things. During each class session, instructors can keep a running list—in smaller classes mentally and in larger classes on a piece of paper—of those students who have contributed to the discussion that day, such as by answering or asking a question. When the same students attempt to volunteer for the second, third, or
subsequent times, instructors can explicitly invite participation from other students, using language such as “I know that there are lots of good ideas on this in here, and I’d like to hear from some members of our community who I haven’t heard from yet today.” At this juncture, wait time is key, as it will likely take time for those students who have not yet participated to gather the courage to join the conversation. If there are still no volunteers after the instructor practices wait time, it may be time to insert a pair discussion, using language such as “We cannot go on until we hear ideas from more members of our scientific community. So, take one minute to check in with a neighbor and gather your thoughts about what you would say to a scientific colleague who had asked you the same question that I’m asking in class right now.” At this point it is essential not to resort to the usual student volunteers and not to simply go on with class, because students will learn from that behavior by the instructor that participation of all students will not be demanded.

11. Learn or Have Access to Students’ Names
For cultivating a welcoming, inclusive, and equitable classroom environment, one of the simplest strategies an instructor can use is to structure ways to get to know and call students by their names. Some instructors may plead an inability to remember names; however, there are many simple ways to scaffold the use of individual student names in a classroom without memorizing all of them. Having students submit index cards with their names and personal information, as described above, is an easy first step to learning names. Additionally, requiring students to purchase and always bring to class a manila file folder with their first names written on both sides in large block letters is another simple way to begin to make students’ names public, both for the instructor and for other students. Instructors who use such folders request that students raise this folder above themselves when asking or answering a question in class, so the instructor can call them by name. More advanced would be for the instructor to personally make the student name tents, preparing perhaps a colorful piece of heavy card stock folded in half, then writing each student’s name in large block letters on each side. The simple act of making the name tags—which is feasible in class sizes of up to 100 students—may aid an instructor in beginning the process of learning students’ names. Regardless of who makes them, these name tents can be tools for a variety of classroom purposes: to call on students by name during class discussions, to encourage students to know one another and form study groups, and to verify names and faces when collecting exams on exam days. In smaller classes, name tents can be used more extensively, for example, by collecting them at the end of class and sorting them to identify members of small groups for work in the next class session. In fact, the attempt to get to know students’ names, and the message it sends about the importance of students in the course, may be more important than actually being able to call students by name each time you see them.

12. Integrate Culturally Diverse and Relevant Examples
Part of building an inclusive biology learning community is for students to feel that multiple perspectives and cultures are represented in the biology they are studying. Although it is not possible to represent aspects of all students’ lives or the cultural background of each student in your course, careful attention to integrating culturally diverse and personally relevant connections to biology can demonstrate for students that diverse perspectives are valued in your biology classroom (Ladson-Billings, 1995). Most topics in biology can be connected in some way to the lived experiences of students, such as connecting what can be an abstract process of how genes produce traits to the very real and immediate example of cancer. Similarly, including examples that connect biology concepts that students are learning to different cultural communities—including both well-known stories like that of Henrietta Lacks and her connection to cell biology and smaller stories like that of Cynthia Lucero and her connection to osmosis—demonstrate to students that you as an instructor want to help them see themselves within the discipline of biology (Chamany, 2006; Chamany et al., 2008). Finally, stories from both the history of science and present-day discoveries, when judiciously chosen, can convey that diverse populations of people can make key contributions in science (e.g., Brady, 2007). Value for the inclusion of diverse perspectives can also manifest in simply being explicit that much of the history of biology has not included diverse voices and that you as the instructor expect this generation of students to literally change the face of the biological sciences.

13. Work in Stations or Small Groups
To promote an inclusive community within the classroom, instructors can integrate opportunities for students to work in small groups during time spent within the larger class. For some students, participation in a whole-group conversation may be a persistently daunting experience. However, instructors can structure opportunities for such students to practice thinking and talking about biology by regularly engaging students in tasks that require students to work together in small...
groups. Care must be taken to be explicit with students about the goal of the group work and, whenever possible, to assign roles so that no student in a small group is left out (Johnson et al., 1991, 1993, 1998; Tanner et al., 2003). It can be challenging to design group work that is sufficiently complex so as to require the participation of all group members. Keeping group sizes as small as possible, no more than three or four students, can mitigate potential for unfairness caused by the act of putting students into groups. As one example, groups of students can be charged to bring expertise on a particular topic to class, check that expertise with others studying the same topic in a small group, and then be “jigsawed” into a new small group in which expertise from different topics can be shared (Clarke, 1994). Additionally, explicit statements from the instructor about expectations that group members will include and support one another in their work can be especially helpful. Finally, in smaller class sizes, an instructor can thoughtfully construct student groups so as to minimize isolating students of particular backgrounds (e.g., attempt to have more than one female or more than one student of color in a group) or interaction styles (e.g., attempt to place quieter students together so that they are likely to have more opportunity to talk). How instructors structure small-group interactions has the potential to provide a feeling of inclusion, community, and collaboration for students who may otherwise feel isolated in a biology classroom.

14. Use Varied Active-Learning Strategies
To engage the broadest population of students, instructors may be best served by using a variety of active-learning strategies from class session to class session. For each strategy, some students will be out of their comfort zones, and other students will be in their comfort zones. Students who may be more reflective in their learning may be most comfortable during reflective writing or thinking about a clicker question. Other students may prefer learning by talking with peers after a clicker question or in a whole class conversation. Still others may prefer the opportunity to evaluate animations and videos or represent their understanding of biology in more visual ways through drawing, concept mapping, or diagramming. One might ask which of these different strategies is the most effective way to teach a given topic; yet this question belies the likely importance of variations in the efficacy of different strategies with different students. There may not ever be a “best” way to teach a particular concept, given the diversity of students in any given classroom. The “best” way to teach equitably—providing access to biology for the largest number of students—may be to consistently provide multiple entry points into the conceptual material for students. The role of an instructor in creating an equitable learning environment that is accessible to all students is to make sure that no single population of students is always outside their comfort zone. If an instructor chooses a singular teaching approach—always lecturing or always concept mapping, regardless of the nature of the approach—it seems likely that the lack of variation could result in the alienation and exclusion from learning of a subpopulation of students. Additionally, using varied active-learning strategies may be key for individual learners to see a concept from multiple perspectives, make multiple associations between the concept and other ideas, and practice a variety of approaches to exploring that concept. By using varied active-learning strategies for each biological topic explored, instructors can work toward building an inclusive and equitable learning environment for a wide range of students with different approaches to learning.

15. Be Explicit about Promoting Access and Equity for All Students
Perhaps the most powerful teaching strategy in building an inclusive and equitable learning environment is for instructors to be explicit that the triad of access, fairness, and classroom equity is one of their key goals. There need not be substantial time spent on conveying this stance, but explicit statements by the instructor about the importance of diverse perspectives in science can make issues of fairness and equity explicit rather than an implicit. Instructors can share with students why they use the teaching strategies they do, for example, sharing the reasoning behind having students write to allow thinking and processing time for everyone. When an instructor publicly asserts that he or she wants and expects everyone in the classroom to be successful in learning biology, students can leave behind the commonly assumed idea that instructors are attempting to weed out students. Being explicit about one’s goal of cultivating an inclusive, equitable, and fair classroom learning environment reiterates that students and instructors are on the same side, not on somehow opposing sides, of the teaching and learning process.

MONITORING (YOUR OWN AND STUDENTS') BEHAVIOR TO CULTIVATE DIVERGENT BIOLOGICAL THINKING
Science is fundamentally about negotiating models and ideas about how the natural world functions. As such, one might expect that undergraduate biology classrooms would mirror this negotiation and consideration of a variety of ideas about how the biological world might function. However, undergraduate biology classrooms have the reputation, likely deservedly, of being forums in which “right” answers—those already accepted as scientifically accurate—are the currency of conversation and the substrate for instructor–student dialogue. Yet research on learning suggests that inaccurate ideas, confusions, and alternative ideas about how the world works may, in fact, be one of our most powerful tools in the teaching and learning process (there are many publications on this subject, among them Posner et al., 1982; National Research Council, 1999; Taber, 2001; Chi and Roscoe, 2002; DiSessa, 2002; Coley and Tanner, 2012). As such, it is important for instructors to cultivate discussion of divergent ideas in classroom conversations about biology—some of which may not be supported by current scientific evidence—as part of the process of moving students toward thinking in more scientifically accurate ways. Given the reputation of science courses as environments in which only those with correct answers are rewarded, biology instructors face the extra and very real challenge of gaining the trust of students to share divergent perspectives. Instructors can begin to establish a classroom community that values divergent ideas and promotes participation by students who may not already have scientifically accurate understanding by using the following four teaching strategies.
16. Ask Open-Ended Questions

One critical tool for instructors aspiring to cultivate divergent biological thinking in their classrooms is the use of open-ended questions, which are those questions that cannot be answered with a simple “yes” or “no” or even easily answered with a single word or phrase. Open-ended questions are by definition those which have multiple possible responses, such that inviting answers from a large group can yield more than an expected set of responses (Bloom et al., 1956; Allen and Tanner, 2002; Crowe et al., 2008). Open-ended questions can be posed orally to frame a class discussion and followed by a quick write or pair discussion to give students time to consider their responses. Alternatively, instructors can plan these questions in advance, so they can be given as brief homework assignments, allowing students time to consider the questions before coming to class. In general, open-ended questions require some design time and may not be easily improvised by most biology instructors. As research scientists, many of us have been trained to ask closed-ended questions, namely questions that drive an experimental design to either confirm or refute a stated hypothesis. In some ways, training in asking closed-ended, experimental questions may be at odds with developing skills in open-ended questioning.

Prior to asking open-ended questions, instructors can attempt to anticipate the likely responses they may get from students. This serves the dual purpose of checking that the question is really all that open-ended, as well as preparing for how one will handle students sharing a wide variety of ideas, which may or may not be scientifically accurate.

17. Do Not Judge Responses

Undergraduate science classrooms in general have the reputation of being places in which only right answers are valued, and participation in class discussions has a competitive tone (Seymour and Hewitt, 2010). However, as instructors, we have the power to encourage all students—not just those who have already constructed biologically accurate ideas—to exercise their voices in our undergraduate biology courses and to make their thinking about biology visible. To create a safe environment that encourages students to share all of their ideas, instructors may be best served in acknowledging student responses as neutrally as possible. This does not require inadvertently supporting a scientifically inaccurate idea. Clearly stating “I’d like to hear from a number of us about our thinking on this, and then we can sort out what we are sure of and what we are confused about,” sets the stage that all the responses may not be correct. Even the most simple “Thanks for sharing your ideas” after each student responds, without any immediate judgment on the correctness of the comments, can set a culture of sharing. It only takes one student experiencing ridicule from a fellow student to immediately bring a halt other students sharing their ideas in class. When such incidents occur, and they will, a simple reminder of the classroom norms, “That is totally not how it works!” to immediately bring those students made to feel uncomfortable can go a long way. Simply using language like, “Could you please keep sharing your ideas? I have no doubt that if you are thinking along these lines, lots of smart people would think that way, too.” Establishing early and regularly enforcing a supportive classroom culture—just as you would in an effective and productive research lab meeting, study section, or any other gathering of scientists—is essential to maintaining an equitable, inclusive, and welcome classroom community.

18. Use Praise with Caution

For instructors new to actively engaging students during class time, or even for seasoned instructors in the first few weeks of a term, it can be challenging to cultivate student participation in whole-group discussions. In response to those students who do share, instructors can unwittingly work against themselves by heaping praise on participating students. “Fabulous answer!” “Exactly!” “That’s perfect!” With very few syllables spent, instructors may inadvertently convey to the rest of the students who are not participating that the response given was so wonderful that it is impossible to build on or exceed. Additionally, in a short period of time, the few students who are willing to participate early in a discussion or the course will become high status in the classroom, those who have reaped the instructors’ praise. Research from sociologist Elizabeth Cohen and her colleagues, described as “complex instruction,” has explored the power instructors have of effectively assigning academic status to students simply by the nature and enthusiasm of their remarks about those students’ responses (Cohen, 1994). So, does this mean instructors should never praise student responses? No. However, it suggests using praise with caution is essential, so other students feel that they still have something to add and can be successful in sharing.

19. Establish Classroom Community Norms

As instructors strive to cultivate a classroom in which divergent and not always scientifically accurate ideas are shared, it is critical that the instructor also establish a set of classroom community norms. In this case, “norms” refers to a set of accepted usual, typical, standard acceptable behaviors in the classroom. Common group norms established by experienced instructors include the following: “Everyone here has something to learn.” “Everyone here is expected to support their colleagues in identifying and clarifying their confusions about biology.” “All ideas shared during class will be treated respectfully.” For many instructors, these classroom norms are simply verbally asserted from the first few days of a class and then regularly reiterated as the term progresses. Importantly, students will observe directly whether the instructor enforces the stated group norms and will behave accordingly. As such, it is important to decide what norms you are comfortable enforcing as the instructor in charge of your classroom. It only takes one student experiencing ridicule from a fellow student based on what they shared (someone shouts out, “That is totally not how it works!”) to immediately bring a halt to other students sharing their ideas in class. When such incidents occur, and they will, a simple reminder of the group norms and public reassurance and support for the student made to feel uncomfortable can go a long way. Simply using language like, “Could you please keep sharing your ideas? I have no doubt that if you are thinking along these lines, lots of smart people would think that way, too.” Establishing early and regularly enforcing a supportive classroom culture—just as you would in an effective and productive research lab meeting, study section, or any other gathering of scientists—is essential to maintaining an equitable, inclusive, and welcome classroom community.
TEACHING ALL THE STUDENTS IN YOUR CLASSROOM

As asserted above, perhaps the most underappreciated variables in teaching and learning are the students themselves and all their individual variations. Although it may be tempting to generalize what students will be like from semester to semester, from course to course, and from institution to institution, there is little evidence to support these generalizations. To promote student engagement and strive for classroom equity, it is essential to constantly and iteratively attend to who exactly is in your classroom trying to learn biology. Below are two specific strategies to help keep the focus of your teaching on the actual students who are currently enrolled in the course you are teaching.

20. Teach Them from the Moment They Arrive

As biology instructors, we assume that the only thing being learned in our classrooms is biology. However, student learning does not begin and end with the biology being explored and discussed. Increasingly, research from a host of fields—educational psychology, sociology, and science education—suggests that learning is not discrete and delimited by concepts under study, but rather continuous and pervasive. Learning is happening about everything going on in the classroom. As such, instructors are best served by considering what students are learning, not just about the subject matter, but also about culture of the classroom from the moment they enter the room. Consider students’ opportunities to learn about classroom culture in just two of many ways: students’ impression on the first day of class and students’ impressions as they enter the classroom for each class session. What an instructor chooses to do on the first day of a course likely sends a strong message to students about the goals of the course, the role of the instructor, and the role of the students. If one wants to convey to students that the course is about learning biology, then reading the syllabus and spending the first class session discussing how grades are assigned is incongruous. Without intent, this instructor may not be the conscious intention of the instructor. Similarly, in inaccessible or too busy to be approached, even though this may not be the conscious intention of the instructor. Similarly, students will likely notice whether the instructor regularly speaks to the same subset of students prior to class each day. In all these cases, instructors can make conscious efforts to convey their interest in and commitment to the learning of all students in the course all the time—before class, during class, and after class, via email. If we want to teach them about biology, we likely need to be teaching them about the culture of our classrooms and their role in it at the same time.

21. Collect Assessment Evidence from Every Student, Every Class

To accomplish the goal of teaching those actual students who are sitting in front of you, it is essential to maximize the flow of information from individual students to the instructor. Frequent collection of assessment evidence—about students’ biological ideas, about their reflections on their learning, about their struggles in the course—is essential for instructors to know the learners they are trying to teach. Beginning immediately, instructors can start with an online “More about You” survey as homework on the first day of a course and continue to collect information about students throughout the semester (Tanner, 2011). For many instructors, this is most easily accomplished through student online submission of writing assignments. Other options include the use of daily minute papers or index cards, clickers, and a variety of other assessment tools (Angelo and Cross, 1993; Huba and Freed, 2000). While the nature of the assessment evidence may vary from class session to class session, the evidence collected from each and every student in a course can aid instructors in continuously re-evaluating student ideas and iteratively changing the arc of the course to best support the learning of that course’s student population. The goal is to assure a constant stream of information from student to instructor, and for each and every student, not just those confident enough to speak up publicly during class. Regular consideration of classroom evidence is foundational for bringing our scientific skills to bear on our teaching.

CONCLUSION

As instructors, we have the power in our classrooms to choose to attend explicitly to issues of access, inclusiveness, fairness, and equity. The strategies presented above are merely starting points from which instructors can step up their attempts to cultivate equitable classroom environments that promote student engagement and participation in learning biology. No doubt this list of equitable teaching strategies could be much longer, and readers are encouraged to record additions that they discover or invent themselves that address the goal of promoting equity and access for all the students in our biology classrooms.

REFERENCES

Campuses, Needham Heights, MA: Allyn and Bacon.

Huba ME, Freed JE (2000). Learner-Centered Assessment on College

Six-thousand-student survey of mechanics test data for introductory
Physics courses. Am J Phys

Hake RR (1998). Interactive-engagement vs traditional methods: a

Haak DC, HilleRisLambers J, Pitre E, Freeman S (2011). Increased
structure and active learning reduce the achievement gap in intro-

Hake RR (1998). Interactive-engagement vs traditional methods: a
six-thousand-student survey of mechanics test data for introductory


Huba ME, Freed JF (2000). Learner-Centered Assessment on College
Campuses, Needham Heights, MA: Allyn and Bacon.


returns to college: what evidence is there that it works? Change 30,
26–35.

Johnson DW, Johnson RT, Smith KA (1991). Active Learning: Coop-
eration in the College Classroom, Edina, MN: Interaction Book.

Labov JB, Reid AH, Yamamoto KR (2010). Integrated biology and
undergraduate science education: a new biology education for the

Ladson-Billings G (1995). But that’s just good teaching! The case for
culturally relevant pedagogy. Theory Pract 34, 159–165.

Lyman F (1981). The responsive classroom discussion: the inclusion
of all students. In: Mainstreaming Digest, ed. AS Anderson, College
Park: University of Maryland.

Moss-Racusin CA, Dovidio JF, Brescoll VL, Graham MJ, Handelsman
Proc Natl Acad Sci USA 109, 16474–16479.

National Research Council (1999). How People Learn: Brain, Mind,

Posner GJ, Strike KA, Hewson PW, Gertzog WA (1982). Accommoda-
tion of a scientific conception: towards a theory of conceptual change.

13.

Rowe MB (1974). Wait-time and rewards as instructional variables,
their influence in language, logic and fate control. Part 1: Wait time.

216.

Rowe MB (1987). Wait time: slowing down may be a way of speeding
up. Am Educator 11, 38–43, 47.

Schinske JN (2011). Taming the testing/grading cycle in lecture
classes centered around open-ended assessment. J Coll Sci Teach 40,
46–52.

Seymour E, Hewitt NM (1997). Talking about Leaving: Why Under-
graduates Leave the Sciences, Boulder, CO: Westview.

Smith MK, Wood WB, Adams WK, Wieman C, Knight JK, Guild N,
Su TT (2009). Why peer discussion improves student performance

Steele CM (1999). Thin ice: stereotype threat and black college stu-

811.

Taber KS (2001). Shifting sands: a case study of conceptual develop-
ment as competition between alternative conceptions. Int J Sci Educ
23, 731–753.

Tanner KD (2009). Talking to learn: why biology students should be
talking in classrooms and how to make it happen. CBE Life Sci Educ
8, 89–94.

Tanner KD (2011). Moving theory into practice: a reflection on teach-
ing a large introductory biology course for majors. CBE Life Sci Educ
10, 113–122.

Cell Biol Educ 1, 3–5.

Tanner KD, Chatman LC, Allen DE (2003). Cooperative learning in
the science classroom: beyond students working in groups. Cell Biol
 Educ 2, 1–5.

Tobias S (1990). They’re not dumb. They’re different. A new tier of


15 June 2013).
Barriers to Faculty Pedagogical Change: Lack of Training, Time, Incentives, and... Tensions with Professional Identity?

Sara E. Brownell* and Kimberly D. Tanner†

*Department of Biology, Stanford University, Stanford, CA 94305–5020; †Department of Biology, San Francisco State University, San Francisco, CA 94132

The time has come for all biology faculty, particularly those who teach undergraduates, to develop a coordinated and sustainable plan for implementing sound principles of teaching and learning to improve the quality of undergraduate biology education nationwide. (Vision and Change, 2011, xv)

Recent calls for reform, such as Vision and Change: A Call to Action, have described a vision to transform undergraduate biology education and have noted the need for faculty to promote this change toward a more iterative and evidence-based approach to teaching (American Association for the Advancement of Science [AAAS], 2011). A key challenge is convincing many faculty—not just a handful of faculty scattered across the country but the majority of life sciences faculty in every institution—to change the way they teach.

Few would disagree that this is an ambitious goal. Change is difficult in any setting, but changing academic teaching appears to be especially tricky. Calls for change imply that the pedagogical approaches our own professors and mentors modeled and taught us might not be the best way to engage large numbers of diverse populations of undergraduates in our discipline. This effort potentially also involves telling faculty that what they have been doing for the past 5, 10, or even 30 yr may not the most effective approach, especially for today’s students. Widespread change in undergraduate biology teaching—or in any of the sciences for that matter—has been documented to be difficult (Henderson et al., 2011).

The general perception is that while there are pockets of change driven by individual faculty, there is little evidence that the majority of our faculty members are reconsidering their approach to teaching, despite dozens of formal policy documents calling for reform, hundreds of biology education research publications on the subject, and the availability and award of substantial amounts of external grant funding to stimulate change toward evidence-based teaching (Tagg, 2012).

In fact, it is somewhat perplexing that we as scientists are resistant to such change. We are well trained in how to approach problems analytically, collect data, make interpretations, form conclusions, and then revise our experimental hypotheses and protocols accordingly. If we are experts at making evidence-based decisions in our experimental laboratories, then what forces are at play that impede us from adopting equally iterative and evidence-based approaches to teaching in our classrooms? What can we—as members of a community of biologists dedicated to promoting scholarly biology teaching—do to identify and remove barriers that may be impeding widespread change in faculty approaches to teaching?

A substantial body of literature has highlighted many factors that impede faculty change, the most common of which are a lack of training, time, and incentives. However, there may be other barriers—unacknowledged and unexamined barriers—that might prove to be equally important. In particular, the tensions between a scientist’s professional identity and the call for faculty pedagogical change are rarely, if ever, raised as a key impediment to widespread biology education reform. In this article, we propose that scientists’ professional identities—how they view themselves and their work in the context of their discipline and how they define their professional status—may be an invisible and underappreciated barrier to undergraduate science teaching reform, one that is not often discussed, because very few of us reflect upon our professional identity and the factors that influence it. Our primary goal in this article is to raise the following question: Will addressing training, time, and incentives be sufficient...
to achieve widespread pedagogical change in undergraduate biology education, or will modifying our professional identity also be necessary?

FOCUSING ON THE BIG THREE: LACK OF TRAINING, TIME, AND INCENTIVES

Insufficient training, time, and incentives are among the most commonly cited barriers for faculty change, and the focus of most of the current efforts to understand and promote faculty pedagogical change (Henderson et al., 2010, 2011; AAAS, 2011; Faculty Institutes for Reforming Science Teaching [FIRST IV], 2012; National Academies of Science/Howard Hughes Medical Institute [NAS/HHMI], 2012).

In terms of training, many faculty have indicated they feel ill-equipped to change the way they teach and thus would like access to structured, formal training. Unsurprisingly, we as faculty may not be knowledgeable about what constitutes a student-centered classroom (Hativa, 1995; Miller et al., 2000; Winter et al., 2001; Hanson and Moser, 2003; Luft et al., 2004; Yarnall et al., 2007) or we may be unconvinced as to whether new teaching methods are really more effective than traditional instruction (Van Driel et al., 1997; Miller et al., 2000; Winter et al., 2001; Yarnall et al., 2007). Even if faculty are aware of reform efforts, science faculty will most likely not have had training in these types of teaching methods (Rushin et al., 1997; Handlesman et al., 2004; Ebert-May et al., 2011). Vision and Change specifically highlights the need for training of early-career scientists, including postdoctoral fellows and assistant professors (AAAS, 2011). Efforts such as the NSF-funded FIRST IV program and the NAS/HHMI Summer Institutes for Undergraduate Biology Education are examples of programs intended to provide postdoctoral scholars and faculty of all ranks, respectively, with the needed expertise in innovative teaching through hands-on training (FIRST IV, 2012; NAS/HHMI, 2012). Although it is too early to gauge the long-term success of these programs, one wonders whether some of these training efforts may be hindered by the lack of buy-in from the home institutions. After faculty go to nationally or regionally organized training workshops and become excited about implementing new teaching strategies, are they met with support or resistance from their colleagues upon return to their home institutions? Furthermore, trying to achieve pedagogical change through 1-d or even 1-wk training sessions seems incongruent with the notion that pedagogical change for any instructor is an iterative and ongoing process. Even the most well intentioned of us forget what we learned, need extra practice, and often revert to our old habits when we are, inevitably, pressed for time. So although it is necessary to provide scientists with training opportunities demonstrating new ways of teaching, training alone is likely insufficient by itself to achieve lasting pedagogical change.

What about issues of time? With the often-competing demands of research and teaching, faculty often find it difficult to carve out sufficient time to reflect deeply upon their teaching. While faculty at different types of institutions have varying degrees of teaching responsibilities, faculty at most 4-yr institutions are also required to do research and obtain significant external grant funding. Although this expectation is most explicit at R1 research institutions, it also exists at many comprehensive institutions, and even at small liberal arts colleges. Regardless of current faculty teaching loads, there is no doubt that the process of changing an instructional technique is time- and labor-intensive (Krockover et al., 2002; Howland and Wedman, 2004; Stevenson et al., 2005; Schneider and Pickett, 2006; Malicky et al., 2007). Additionally, research has shown that interactive teaching, as compared with traditional lecturing, typically takes more preparation time (Miller et al., 2000; Hanson and Moser, 2003; Pundak and Rozner, 2008). Thus, not only will the actual process of change take more time, but we are asking faculty to shift to a method that might be, by its very nature, more time-consuming. Institutional recognition of this fact, and corresponding allowance in faculty schedules, will thus be critical to accomplishing widespread adoption of evidence-based teaching strategies. In addition, for such changes to be made, there needs to be an incentive for faculty to modify their pedagogical approach; even though time is necessary, time alone is likely not sufficient for widespread change to occur.

Incentives likely drive most of our professional decisions, and teaching is no exception. If we as faculty are indeed provided the requisite training and time to enact changes in our teaching, then there must also be a concomitant reason why we should want to change. Research has demonstrated that even if faculty are interested in changing their pedagogical approach, few incentives are available to spur this action (Hativa, 1995; Walczyk and Ramsey, 2003; Gibbs and Coffey, 2004; Weiss et al., 2004; Wilson, 2010; Anderson et al., 2011). Many argue that if change takes time and training, then faculty need to be compensated for their efforts in the form of lower teaching loads, financial benefits, recognition for tenure, teaching awards, or even, at the most basic level, verbal acknowledgment from colleagues and supervisors. Research has shown that in many universities there are few to no rewards for teaching in novel ways or introducing evidence-based strategies (Kember and McKay, 1996; Frayer, 1999; Krockover et al., 2002; Romano et al., 2004). In fact, there are some reports that change in instruction can lead to poor teaching evaluations, due to student resistance to change, which can negatively affect progression to tenure (Anderson, 2002, 2007). Until universities reward teaching as much as research (Hannan, 2005; Porter et al., 2006) or find ways to better integrate teaching and research (Kloser et al., 2011), the pressure is on faculty, in particular pretenure faculty, to spend the majority of their time on research, sometimes at the expense of high-quality teaching or any attention to the constant calls for change in teaching practice.

The needs for training, time, and incentives are the most commonly cited impediments to widespread change in undergraduate biology faculty teaching practice, and indeed these are real and present barriers. However, let us pause. Imagine a university that provides faculty with all the training, all the time, and all the incentives faculty needed—would that be enough for all biology faculty or even the majority of biology faculty to adopt or build on pedagogical reform? While these “big three” factors are likely necessary for change to occur, it is far from clear that they are sufficient for it to happen. Focusing our efforts exclusively on training, time, and incentives ignores at least one additional and potentially key barrier to faculty change that is largely absent from change discussions: the role of a scientist’s professional identity.
INTRODUCING THE CONCEPT OF A SCIENTIST’S PROFESSIONAL IDENTITY

The process by which we become scientists is often so long and arduous that few of us may have actually taken the time to reflect what constitutes our professional identities as scientists. In the midst of mastering laboratory techniques and crafting research grants, we are also learning, often subconsciously and implicitly, what professional norms we need to obey, or at least tolerate, to be perceived as successful academic scientists.

Identity is most often thought about in the social sciences in terms of personal identity or how a person thinks of himself or herself in the context of society. Based on the ideas of Mead (1934) and Erikson (1968), identity is not a stagnant property, but rather an entity that changes with time, often going through stages, and is continuously modified based on the surrounding environment. It has been described as “being recognized as a certain kind of person in a given context” (Gee, 2001, p. 99).

For the purposes of this article, we consider scientists’ professional identities to be how they view themselves and their work in the context of their disciplines and how they accrue status among their professional colleagues as academic scientists. These aspects are heavily influenced by the training specific to academic scientists, including course work, laboratory experiences, and the everyday culture and rewards of the scientific profession. Peer acceptance, or more formally the process of peer review, is also closely tied to the development of a professional identity in the sciences. Both the publication of the research we accomplish and garnering the resources we need for experimental work, either at our institution or from national funding agencies, are generally dependent on positive peer review and a shared professional identity with these peers.

Thus, the development of a professional identity is not unlike the development of a personal identity but is situated in the context of a discipline and thus framed by the “rules of membership” of that discipline. If you are an academic scientist, then it is likely you were either explicitly told the rules of academic science, or you were able to somehow infer them and make choices to fit in or at least make others think that you fit in. Frustratingly, these rules of professional membership are not always obvious or intuitive, sometimes inadvertently keeping out those who are not afforded opportunities to learn the rules, expectations, and currencies of status within a particular discipline. This has been previously documented as a pivotal problem in the sciences, in particular in attracting and retaining women and people of color in the field (Carlone and Johnson, 2007; Johnson, 2007).

While a professional identity is by definition an internalized identity, it guides our external actions and decisions in our profession, including the decisions we make about how we teach. If a scientist has a professional identity that does not encompass teaching at all, or if a scientist has a professional identity he or she feels could be put at risk in his or her discipline and among his or her peers by embracing innovative approaches to teaching, then professional identity becomes a critical barrier in efforts to promote widespread change in undergraduate biology education.

WHAT ARE THE TENSION POINTS BETWEEN MAINTAINING ONE’S SCIENTIFIC PROFESSIONAL IDENTITY AND PARTICIPATING IN PEDAGOGICAL CHANGE?

Several lines of inquiry support why a scientist’s professional identity might interfere with his or her willingness to participate in pedagogical change. We describe here three tension points that individual faculty may commonly encounter when deciding whether or not to participate in biology education change efforts: 1) training cultivates a primarily research identity and not a teaching identity, 2) scientists are afraid to “come out” as teachers, and 3) the professional culture of science considers teaching to be lower status than research and positions scientists to have to choose between research and teaching. Each of these tension points, along with research literature that explores its origins, is presented below.

TRAINING CULTIVATES PRIMARILY A RESEARCH IDENTITY AND NOT A TEACHING IDENTITY

The first tension point between professional identity and pedagogical change efforts is that scientists are trained in an atmosphere that defines their professional identities primarily as research identities to the exclusion of teaching identities. A scientist’s professional identity is shaped by a number of factors, but this socialization into the discipline of science often begins in graduate school (Austin, 2002). For undergraduates who spend considerable time in research labs for summer research projects or honors theses, socialization may begin earlier. However, graduate school is when all future scientists formally enter a learning period about the scientific profession and the cultural norms of the profession, often leading aspiring young scientists to adopt the values, attitudes, and professional identities of the scientists who trained them. Graduate school is the shared playground, where scientists learn the culture and values of the field, as well as how to play the game of professional science.

Over the past 30 yr, doctoral and postdoctoral training at research institutions has put a tremendous emphasis on research, immersing students in the culture of research for a scientific discipline, while often ignoring teaching (Fairweather et al., 1996; Boyer Commission on Educating Undergraduates in the Research University, 2002). While some time spent as a teaching assistant may be required, in general there is no requirement for evidence of developing competency in teaching. Consequently, it has been asserted that there is a profound disconnect between the training that students are receiving in doctoral programs and the careers that many of these students will ultimately enter (Tilghman, 1998; Golde and Dore, 2001; Austin, 2002; Dillenburg, 2005; Dillenburg and Connolly, 2005; Fuhrmann et al., 2011). Faculty positions at most colleges and universities are primarily teaching positions, and even faculty positions at research institutions require some teaching, but the majority of graduate students in the sciences are only taught how to do research.

What support is given to those graduate students who are interested in developing teaching skills in graduate school? A growing number of institutions have graduate student and
faculty teacher-training programs (Rushin et al., 1997; Austin et al., 2008; Ebert-May et al., 2011). However, despite recommendations for the implementation of pedagogy-focused training in graduate school, programs focused on innovative teaching strategies are often voluntary and serve only a small percentage of the overall population of graduate students. Currently, there are no federal mandates associated with training grants that would require pedagogical training for future scientists.

As a result, most graduate students still learn how to teach through an “apprenticeship of observation” (Lortie, 1975; Borg, 2004). They model their own teaching approaches after their professors. Students without explicit training tend to teach “naively” (Cross, 1990), often relying on inaccurate assumptions about teaching and learning. Most college classes in the sciences are taught in the traditional lecture format, so the majority of beginning science instructors equate teaching with lecturing, both linguistically and conceptually (Mazur, 2009). Without explicit training during graduate school, postdoctoral training experiences, or even early faculty years, these inaccurate assumptions about teaching appear to persist and become solidified. Additionally, even if a scientist trainee or early-career faculty member is interested in adopting pedagogical approaches different than the norm, there may be peer pressure from scientific colleagues to conform to traditional methods of teaching (Van Driel et al., 1997; Gibbs and Coffey, 2004).

Not only is teaching not a formal or recommended component of postdoctoral training, some faculty advisors even view teaching as completely ancillary to, and a distraction from, the training that postdoctoral scholars need, ostensibly to become professors. The National Institutes of Health’s Institutional Research and Academic Career Development Awards (NIH IRACDA) postdoctoral program is a notable exception to this. IRACDA postdoctoral fellows conduct research in basic science at R1 institutions and concurrently have formal, mentored teaching experiences at minority-serving institutions (IRACDA, 2012); however, IRACDA currently serves only a limited number of postdocs. Additionally, the FIRST IV program also seeks to provide postdoctoral fellows with training and mentored teaching experiences as they transition to faculty roles, but again, this is an option for a limited number of postdocs (FIRST IV, 2012). Both of these programs could serve as models for the more widespread integration of teaching and research into the scientific training and professional identity development of postdoctoral fellows. If scientists do not consider teaching part of their professional identities, then how can we expect them to change their own teaching and, even more importantly, support and encourage others to change as well?

SCIENTISTS ARE AFRAID TO “COME OUT” AS TEACHERS

A second tension point between maintaining one’s professional identity and participating in pedagogical change is that embracing a teaching identity as part of one’s scientific professional identity can be perceived as a liability and something to be hidden. Mark Connolly and colleagues have documented that some graduate students who are interested in teaching are afraid to “come out” as teachers (Connolly, 2010). They fear that they will be marginalized and discriminated against by their scientific peers and mentors. Some faculty advise graduate students to hide their interest in teaching; these mentors worry that the rest of academia will not take such students seriously as researchers (Connolly, 2010). There have been reports that some research professors, upon learning their graduate students are interested in teaching, no longer spend the same amount of time mentoring them. Significantly, some doctoral students have faculty advisors who do not allow them to engage in any activities outside laboratory work (Wulff et al., 2004). Some advisors are of the mentality that graduate students should always be at the bench and that any time devoted to teaching negatively affects research, despite a recent study indicating that teaching while doing research might improve research skills (Feldon et al., 2011). Unfortunately, this approach leaves students with both a skill set and perspective on science that is very narrowly focused. Postdoctoral scholars often face similar problems but often without the larger support structure that many graduate students have. Because postdocs tend to be fairly isolated in individual labs, they are even more dependent on their research mentors for guidance about career paths.

If graduate students and postdoctoral scholars fear the ramifications of admitting that teaching is part of their identity, an interest in teaching can be internalized as something illicit, to be kept hidden from peers and mentors. Even those who are interested in continuing in academia to become professors are encouraged to limit the amount of teaching they do. This implicit, if not explicit, research-centric norm of graduate school can result in a student’s internal conflict between developing a professional identity as a research scientist and a desire to also develop part of a professional identity as a teacher. As students struggle to reconcile these aspirations, they can fall prey to believing that teaching is inherently inferior to research and that if they are to succeed in the academic world of science, they should focus exclusively on research. For a graduate student with a strong interest in teaching, this could even result in doubts about his or her ability as a scientist. In the process of embracing a teaching identity, budding scientists potentially risk their status as researchers, as well as their professional identities, status, and even membership within the scientific community.

THE PROFESSIONAL CULTURE OF SCIENCE CONSIDERS TEACHING TO BE LOWER STATUS THAN RESEARCH AND POSITIONS SCIENTISTS TO HAVE TO CHOOSE BETWEEN RESEARCH AND TEACHING

Finally, a third tension point between maintaining one’s professional identity and participating in pedagogical change is that teaching is often regarded as lower status than research in the scientific disciplines (Beath et al., 2012). A large part of this disparity in status originates from the culture of individual laboratories, departments, institutions, and even the discipline as a whole (Cox, 1995; Quinlan and Akerlind, 2000; Marbach-Ad et al., 2007). However, it is also reinforced by the general salary and status structures with regard to teaching within our society, in which teaching is generally considered to be not as well compensated for or afforded as much respect as many other professions.

Faculty members who want to be perceived as successful and “real” scientists may have purposely avoided integrating
BRINGING PROFESSIONAL IDENTITY TO THE FOREFRONT OF CHANGE DISCUSSIONS: SHIFTING FROM AN INSTITUTIONAL DEFICIT MODEL TO A DISCIPLINE DEFICIT MODEL

Given the tension points described above, professional identity may not be just one additional barrier to faculty pedagogical change; it could be hypothesized to be a key underlying reason why change strategies addressing training, time, and incentives have to date had only limited success in engaging broad groups of faculty in widespread biology education reform. If biology faculty are potentially entrenched in a professional identity grounded in a research identity to the exclusion of a teaching identity, then it would behoove us, as a community, to consider the possibility that professional identity could undercut all our efforts centered on the “big three” change strategies. As a scientist grounded in a research identity, one may view pedagogical training with skepticism, considering it to be a waste of time and effort, in particular if the training tries to promote teaching methods that depart from the cultural teaching norm in science: lecturing. In addition, it follows that extra time might not be the answer to promoting faculty change, if tensions with professional identity are at play. If we have extra time in the day, we may more likely spend that time on research activities that raise our status with professional colleagues and are aligned with our professional identities. Finally, tensions between a professional scientific identity and teaching reform may, unfortunately, trivialize any teaching incentives that are developed. If scientists have professional identities that are predominantly research identities, then a Nature report or Science article will always be viewed as higher status than a departmental, university-wide, or even a national teaching award. Giving incentives for teaching will likely only have positive effects if we, as a scientific community, somehow begin to value those incentives to the same degree as research-based incentives.

A common approach when we think about the reasons why faculty might not change the way they teach is to raise questions about the culture of individual institutions. We assume that the department or institution does not offer training opportunities, release time to develop new courses, or incentives for teaching in scientific ways. This could be broadly classified as an “institutional deficit model,” in which the institution lacks what is needed for reform. Certainly such problems can be inhibiting, and where they exist, institutional reform may be necessary to promote widespread involvement of faculty in pedagogical change. Many of the current pedagogical change strategies and frameworks operate within this model (Henderson et al., 2010, 2011).

However, if we approach the issue of faculty change through the lens of professional identity, we will also want to consider a “discipline deficit model.” Faculty are not only members of their campuses, but also of their national professional societies and the professional community of scholars working in their particular fields. Perhaps it is not only a matter of institutions needing to provide training, time, and incentives, but also a need for a disciplinary culture shift, such that there are both a sufficient level of status attached to teaching and a critical mass of individuals who have professional identities that include teaching. Some might argue that regardless of what institutions offer, most faculty will not change the way they teach, because they view teaching as accessory to their professional identities, derived not from their institutions, but rather from their disciplines, which are cross-institutional.

Finally, there is clearly a need for much more empirical research on all the potential barriers to faculty pedagogical change, but especially on the role of professional identity in determining whether a scientist chooses to participate in biology education reform efforts. Would efforts to broaden the professional identities of scientists to include teaching accelerate pedagogical change? To what extent do graduate or postdoctoral pedagogical training programs alter the professional identities of these early-career scientists? What are the long-term impacts of programs such as FIRST IV, NIH’s IRACDA, or the HHMI/NAS Summer Institutes, in particular in terms of whether participants are more or less likely to engage in pedagogical reform compared with others? How would biologists—with a range of involvement in teaching...
and biology education reform efforts—their professional identities and how these identities shape their professional choices and aspirations?

LOOKING FORWARD: HOW COULD WE ALTER OUR PROFESSIONAL IDENTITIES TO BE MORE INCLUSIVE OF TEACHING?

To achieve widespread pedagogical change toward more iterative and evidence-based approaches, it appears that we need to find ways to challenge the assumption that a scientist’s professional identity should be primarily research-focused and consider ways in which teaching could become more integrated into the fabric of the discipline. Three possible areas for action are explored below.

First, one place to start would be to broaden the goals and content of doctoral and postdoctoral training. Instead of having a handful of unstructured teaching requirements, students could be enrolled in training programs specifically designed to give them mentorship and support to teach in scientific ways. Specific faculty could be identified as teaching mentors for graduate students, who in turn could be given increased teaching opportunities and responsibilities as they progressed through the program. An important caveat is that these teaching mentors would themselves need to be properly trained in scientific teaching. In addition to excellence in research, excellence in teaching would also be an expected outcome of graduate education. One could envision a requirement in which dissertations included a chapter that provided evidence of scholarship and achievement in teaching. Those agencies and foundations that fund graduate education in the life sciences could take the lead in requiring such pedagogical training and deep experiences with teaching for the graduate students they support. By better integrating teaching within the current structure of scientific training, one could provide the next generation of scientists with a better foundation and skill set and also foster a teaching identity as part of their professional identities.

A second way to better align professional identity with the goals of widespread pedagogical change may be to target the place where many faculty derive and maintain their professional identities: scientific journals. Publication and peer review in these journals is an important aspect of professional identity. Some scientific journals are beginning to include education sections, but these are often commentary, rather than research articles. An exception to this is Science magazine, in which a number of education articles have appeared as research reports over the past few years. By including articles about scholarly teaching and education research, scientific journals can influence scientists to view scientific teaching as a part of their professional activities. Notably, a number of scholarly journals that maintain high standards of peer review and national/international distribution have been developed in recent years that provide biologists with a venue for publication of their pedagogical research. CBE—Life Science Education, supported by the American Society for Cell Biology and the HHMI, is a good example of growth in this area. There has been a recent push to integrate peer-reviewed education articles from journals such as CBE-LSE into the tables of contents of scientific journals of professional societies, to provide more faculty easier access to education articles most relevant to their fields. This may enable scientists to view education articles and often by association, teaching, as important characteristics of their professional identities.

Third, a key venue in which scientists construct and maintain their professional identities is at scientific professional meetings. These meetings were generally founded with a research focus, but many professional societies now have education sections within their annual meetings. Unfortunately, these are often not well integrated into the rest of the scientific meeting—sometimes entailing additional costs and being located in different venues and held on different days—reinforcing the concept that the education meeting is distinct from the research meeting. In addition, how are education research findings presented at these conferences? Ironically, the oral presentations are almost always presented as lectures, even when the topic of the talk is about how lecturing is not very effective! This illustrates how prevalent and influential the assumptions are about the expected norms of behavior and interaction at a scientific conference. Even biologists who have strong teaching identities and are well aware of more effective ways to present findings choose, for whatever reason (professional culture? professional identity?), not to employ evidence-based teaching and communication methods in the venue of a scientific conference. And while workshops and poster sessions would allow a higher level of interaction and dialogue—both generally more effective means of conveying information than oral presentations—these venues are often perceived as less important, lower status, and less stringent for high-quality data in the culture of scientific conferences.

IN CONCLUSION . . .

The challenge of addressing tensions between professional identity and pedagogical reform is a complicated issue. Importantly, we need to keep in mind that we as scientists ourselves are the ones responsible for the current state of our professional identities. We as academic scientists set up the tenure structure, publication requirements, and training requirements and dictate the group norms and expected modes of interaction in our own disciplines. We have created and contributed to a culture of science in which research generally has higher status than teaching. Some faculty continue to perpetuate the myth that a researcher should not want to teach and broadcast that value judgment to new graduate students, who are trying to forge their way as scientists. But we, as a professional community, also have the opportunity to take steps to broaden our professional identities and in doing so, address a potentially critical barrier in achieving widespread biology education reform.

REFERENCES


Austin A (2002). Preparing the next generation of faculty: graduate education as socialization to the academic career. J Higher Educ 73, 94–122.


Boyer Commission on Educating Undergraduates in the Research University (2002). Reinventing Undergraduate Education: Three Years after the Boyer Report, Stony Brook, NY: Stony Brook University.


Marbach-Ad G et al. (2007). A faculty team works to create content linkages among various courses to increase meaningful learning of targeted concepts of microbiology. CBE Life Sci Educ 6, 155–162.


Color may be evolution's most beautiful accident. Cristina Luiggi, writing in The Scientist

To understand how life works, it is essential to understand physics and chemistry. Most biologists have a clear notion of where chemistry fits into their life sciences research and teaching. Although we are physical beings, physics does not always find a place in the biology curriculum. Physics informs and enlightens biology in myriad dimensions (Figure 1), yet many biology courses proceed with little or no consideration of physical properties or principles. Phil Nelson, a biophysics professor at the University of Pennsylvania, asserts, "Physical models are often weirdly, unreasonably effective in stripping away the inessential from a biological system—and in displaying connections between things that seemed not obviously connected to our untrained imagination." In this review of online media in the realm of biological physics, I will meander up and down the scale of nature to explore the intersection between physics and biology. Let us begin with the macro scale and the most interesting subject of all, ourselves.

While your cat can likely outrun you in a short- or middle-distance race, you would probably win a longer race. Our species excels at long-distance endurance running, and in fact, running prey to exhaustion is a hunting technique used by wolves and modern humans. Dan Lieberman at Harvard University is a leading researcher in the study of the functional form of the human body. He and his colleagues have a website devoted to barefoot running (http://barefootrunning.fas.harvard.edu), a fad that has pushed the limits on acceptable footwear at the office. How is it that, in our shoeful society, where the average American cannot run a mile wearing shoes, some individuals can run hundreds of miles a month in bare feet? In addition to dodging hazards, a barefoot runner learns to strike the ground first with the forefoot rather than with the heel. Striking with the forefoot dampens the amplitude of high-impact transients that heel strikes generate. The strong impacts generated by heel strikes can also be dampened by the elevated and well-cushioned heels of running shoes, which were first developed in the 1970s (Figure 2). So which is the better way of running? Does one way minimize bodily damage from impact transients better than another? We...
do not know for sure yet, and students might be interested in learning more about the methods that Lieberman and his colleagues are employing to get answers. You can view good videos illustrating their research by clicking on Biomechanics and Videos at the top navigation bar of their website. However, I recommend first viewing a video produced by Nature to go with a Lieberman journal article (link is available at the bottom of this page: http://barefootrunning.fas.harvard.edu/1WhyConsiderFootStrike.html). It features Lieberman himself barefoot running in Massachusetts, on pavement, in the winter. In 6 min, you will see hypotheses and methods of study, and students should be able to see from the force transduction graphs how different the profile is for a heel strike versus a forefoot strike. There are important biological questions wrapped up in the force mechanics. What are the evolutionary trade-offs in adapting bones to specific functions? How can an organism’s power and speed be balanced against mechanical stress and tissue trauma?

While our legs absorb huge impacts when we run, other animals use body parts in even more punishing ways to obtain food, for example, to penetrate the armor of prey. In the pantheon of amazing creatures, the stomatopods, commonly called mantis shrimp (~400 known species), have gained fame thanks to clever research combining physical and biological perspectives. Sheila Patek’s TED talk on her stomatopod research is an excellent introduction to stomatopod biology (www.ted.com/talks/sheila_patek_clocks_the_fastest_animals.html). Stomatopods have the fastest predatory strike in nature; some species use specialized forelimbs to club armored prey with great force, while other species rapidly spear fast-moving prey (Figure 3). One of the most amazing physical phenomena observed in connection with these strikes is light-emitting cavitation. Patek began studying this biologically produced cavitation when she recorded surprisingly large second peaks following the initial force generated when a stomatopod was induced to hit a force transducer (smeared with shrimp paste). She wanted to understand what was causing that second peak, and why it was happening mere milliseconds after the initial strike, a nice example of a physical measurement driving further biological inquiry. It turns out that the stomatopod limb is traveling through the water so fast and striking the prey’s shell so forcefully that water is being vaporized and forming rapidly collapsing bubbles that release huge additional energy in the form of heat, pressure, and light just milliseconds after the initial strike. High-speed video captures the vapor cloud and light flash approximately 13 min into the 16-min TED talk. Cavitation is familiar as a force that destroys high-speed boat propellers, and eventually it destroys stomatopod raptorial limbs as well (physics rules), but the predators just grow a new one (biology rules, too). Analyzing the forces required and energies involved in cavitation would be an excellent biophysics exercise.

Dr. Patek has built a handsome faculty website (www.bio.umass.edu/biology/pateklab/users/sheila-patek), with images, videos, and good explanations of her research interests. Legendary invertebrate neurobiologist Malcolm Burrows was the first to analyze the biomechanics of stomatopod raptorial limbs, and Roy Caldwell at Berkeley has been a leader in probing compelling aspects of stomatopod biology. His website (www.ucmp.berkeley.edu/aquarius/signals.html) is a fascinating window into stomatopod biology, literally, because Dr. Caldwell was able to study stomatopods in their native habitats through submarine windows as part of an Aquarius project. Coral Science, an organization devoted to making coral reef research accessible, also provides a good feature on stomatopod biology (www.coralscience.org/main/articles/reef-species-4/stomatopods).

Some of the best material on the Web is in blogs these days, and a good example is a blog by Marc Srour, a young paleontologist interested mostly in arthropods (http://bioteaching.wordpress.com/2011/07/13/mantis-shrimp-crustacea-stomatopoda). Casey Dunn’s lab at Brown University has a blog project called Creature Cast (http://creaturecast.org/about) that is devoted to telling zoology stories, incorporating contributor videos, mostly from graduate students and postdocs. Students might enjoy seeing the simple animated story on stomatopods (http://creaturecast.org/archives/2054-creaturecast-the-stomatopod-strike).

Sheila Patek’s research has focused on a particular structure of the stomatopod raptorial limb called the saddle. The saddle is a section of shell at the strike pivot joint that is surrounded

---

**Figure 2.** Measurements of the forces occurring at the ground for (a) a barefoot runner vs. (b) a shod runner when striking with the heel. Note how much bigger and sharper the initial impact transient is when not cushioned by a shoe. (From http://barefootrunning.fas.harvard.edu.)
The saddle shape is a member of an elegant family of geometric structures called hyperbolic paraboloids (Figure 1, left). According to Patek, hyperbolic paraboloids are well known to mathematicians and are used extensively in engineering, architecture, and jewelry-making, because maximum strength is provided with minimal material, and the geometry distributes forces evenly across a light structure. (Next time you eat a Pringles potato chip, do so in the name of science education, and note the shape.) The saddle, along with other structural mechanisms, allows the stomatopod limb to store and rapidly release tremendous amounts of shell-crushing energy while using much less muscle mass than otherwise necessary.

Although the raptorial limbs of stomatopods are remarkable, perhaps the most amazing feature of stomatopods is their eyes (Figure 4), and exploring the science of vision presents one of the most engaging ways to combine physics and biology. A 5-min video from the Australian Broadcasting Corporation (www.abc.net.au/catalyst/stories/3280489.htm) is a concise primer on stomatopod vision. This video gives good explanations of why mantis shrimp can see many more colors than we can and can detect linearly and circularly polarized light. They are the only animals known to naturally detect circularly polarized light. We use circularly polarized light while sitting in a movie theater watching a 3D film through glasses with left and right polarizing lenses. The Not Rocket Science blog (http://scienceblogs.com/notrocketscience/2008/03/21/mantis-shrimps-have-a-unique-way-of-seeing), associated with National Geographic, has a good discussion of how mantis shrimp may make use of their remarkable visual capacity. Michael Bok, who also studies the stomatopod visual system, has an excellent blog on arthropods (http://arthropoda.southernfriedscience.com). He has some great underwater color photography, including some photos of larval mantis shrimp.

From childhood, we wonder at the world’s color and ask “Why?” And you may be wondering why the mantis shrimp’s visual system should have such remarkable capabilities. “Why” questions can be tricky. In the case of vision, research can provide deep insights into evolutionary adaptations and stimulate biologists to use physics to better understand biology. It is challenging to understand how stomatopods use their ability to see circularly polarized light for specific adaptive behavior. But from the physics perspective, we can study how the visual system extracts information from the physical world to gain some insight.

NBC news has a science portal (www.nbclearn.com/portal/site/learn/chemistry-now/chemistry-of-color) with a video focused on why flowers are colored. The transcript is available, as are a couple of lessons tied to the video. The video is a case study in how we tend to define the differences among biology, chemistry, and physics as methods of natural science inquiry. The video asserts that flowers are colored because of genes, a biological explanation. These
genes make pigments that are stored in cellular structures, an explanation that mixes biology and chemistry. The video proceeds to physics, presenting the beginnings of a quantum mechanics-based explanation of how color is generated. Asking questions about flower color is an engaging way to unite science, technology, engineering, and mathematics disciplines in a single case study.

From a chemistry–physics viewpoint, there are three types of color in the world: pigment color, structural color, and luminescent color. Each type of color arises from molecules interacting with components of light in completely different ways. The online exhibit Why Things Are Colored (www.webexhibits.org/causesofcolor/7I.html) provides a good context for thinking about how biological colors are produced and perceived. Color from pigment is the most familiar. If you grind pigment, you produce finer and finer grains of the same color. Using mechanical and chemical processes, you can extract the pigment and use it to color other things. Pigment molecules absorb light of specific frequencies and reflect the light of other frequencies. The United Kingdom Chemguide website (www.chemguide.co.uk/inorganic/complexions/colour.html) dives a little deeper into the atomic-level explanation for pigmented color. Chemguide makes it clear that our understanding of color is incomplete from a quantum physics point of view. If you want more in-depth lessons, the Physics Classroom (www.physicsclassroom.com) has excellent materials, including a section on the nature of light and vision (www.physicsclassroom.com/class/light).

In contrast to color from pigment, structural color arises from light interacting with structures at a nanometer scale, the same scale as the wavelengths of light (Figure 5). If you change these nanostructures, you eliminate or alter the color. If you grind the “colored” material, you do not produce smaller bits of color, you destroy it. It could be helpful to think of structural color as a form of nanotechnology. Over time, pigments fade via chemical degradation processes, while structural color can remain amazingly vibrant long after an organism’s death. The Scientist website recently published a nice feature on structural color (www.thescientist.com/?articles.view/articleNo/34200/title/Color-from-Structure). A fascinating aspect of structural color is the diverse ways that plants, birds, and insects achieve structural color. In evolutionary convergence, beetles and birds, for example, decorate themselves blue, using entirely different macro and micro structures, with the common result being that the structures refract and reflect light such that we see blue at the surface. A good explanation of the basics of structural color can be found in a science blog affiliated with The Guardian newspaper (www.guardian.co.uk/science/punctuated-equilibrium/2007/oct/16/birds-physics). Shuichi Kinoshita and Shinya Yoshioka have published a superb review of structural color in nature (http://openwetware.org/images/8/80/Kinoshita_StructColorsNature_ChemPhysChem2005.pdf) and Minoru Taya has posted some useful curriculum materials on structural color in nature from his mechanical engineering course at the University of Washington (http://courses.washington.edu/mengr568/notes/color_in_nature.pdf). For a more mathematical treatment of structural color, have a look at a review by Michael Steindorfer and colleagues in The International Online Journal of Optics (www.opticsinfobase.org/oe/fulltext.cfm?uri=oe-20-19-21485&id=241224). In addition to blues and violets, you might be surprised to learn that white is a common structural color, since brilliant whiteness is very difficult to achieve with pigments alone. In practice, many of the colors we perceive in nature are a combination of pigment and structural color.

Richard Prum and others have been studying structural aspects of bird feathers for some time, examining small air pockets and organized arrays of keratin granules produce structural colors. Recently, Prum and colleagues published an article (www.pnas.org/content/early/2010/06/11/0909616107.full.pdf+html?with-ds=yes) on the structural
Figure 5. Structural color arises not from pigment, but from light being refracted and reflected while interacting with fine-scale structures. The berries are blue from structure, while the squid skin exhibits myriad colors from a combination of pigment and structure. (Berry photo courtesy of P.J. Rudall. Squid photo courtesy of Grayson Hanlon from The Scientist.)

color of butterfly wings, some of the most beautiful palettes known. Butterfly wings are covered with scales with microstructures composed of gyroids (Figure 1). Wired magazine has an online article on gyroids and structural color in butterfly wings (www.wired.com/wiredscience/2010/06/butterfly-colors). For a more technical discussion on butterfly wing gyroids and their diversity across species, see the amazing work of K. Michielsen and D. G. Stavenga (http://rsif.royalsocietypublishing.org/content/5/18/85.full). The gyroid shape is in the family of minimal surfaces, structures that are appealing if you want to make things very strong while keeping them very light. Interestingly, you could say that Alan Schoen discovered the gyroid in the 1970s while working for NASA, and you can learn more about the interesting history of this structure at his website (http://schoengeometry.com/e_tpms.html). Adam Wyhaupt published a concise article on gyroids with good graphics in +plus magazine (http://plus.maths.org/content/meet-gyroid). To have some fun playing around with gyroids and other triply minimal surfaces, visit Paul Nylander’s website (www.bugman123.com/MinimalSurfaces/index.html).

One of the coolest things about how organisms produce color is that the morphology responsible for both pigment color and structural color can be detected in fossils (Figure 6). In the case of pigment, melanosome type and distribution can be used to deduce the color patterns of fossilized birds, mammals, and dinosaurs. In the case of structural color, patterns can be deduced and even seen intact when preservation is so fine that the structural color is retained in the fossil. Discovery News has a short feature on “paleocolor” research (http://news.discovery.com/animals/psychedelic-colored-insects-111115.htm). Derek Briggs at Yale University is a pioneer in paleocolor, his research stemming from his group’s work on remarkably well-preserved fossils from the Cambrian and Ordovician time periods (http://people.earth.yale.edu/profile/deb47/content/research). Jakob Vinther, Briggs’ former student, has an excellent Web page on deducing color from fossils (www.jakobvinther.com/Fossil_color.html).

Interestingly, the nanometer scale of structures that accounts for structural color are smaller than the molecular scale that cellular enzymes control directly. Therefore, structural color results from self-assembly and autonomous processes that arise from the inherent properties of the constituent monomers and conditions in the extracellular milieu. Research on complex, nanoscale biological structures is in its infancy. Rather than list technical publications, I think it is useful to get students started on thinking about self-assembly principles in general. A good way to prime students to think about the rules, conditions, and constraints for self-assembly is to play the protein-folding “game” Fold-it. To get started, visit the Fold-it portal information.

Figure 6. The colors of long-dead organisms can be deduced from exceptionally well-preserved fossils. A fossil feather (left) has a melanosome conformation similar to a modern feather (center). (From Jakob Vinther [www.jakobvinther.com/Fossil_color.html].) On the right, an exceptionally well-preserved fossil beetle carapace shows intact blue structural color. (From Derek Briggs [http://people.earth.yale.edu/profile/deb47/content/research].)
I have barely scratched the surface on the biology–physics interface in this Feature, and I hope readers are intrigued by the science education opportunities afforded by entwining physics and biology. I will close with a few comments relating to biology–physics curricula.

As many readers are aware, there is a decades-old “physics first” educational movement. The original idea was that U.S. high school students should take physics in ninth grade as a foundation for further courses in chemistry and biology. There have also been related movements and discussions at the college level. The move to put physics first in the science curriculum has had its ups and downs, mostly losing gains made in the 1990s but not completely disappearing. You can see an ongoing implementation at A TIME for Physics First (www.physicsfirstmo.org), which is affiliated with the University of Missouri–Columbia. My employer, the Howard Hughes Medical Institute, has helped to fund a project called NEXUS (http://umdberg.pbworks.com/w/page/31612279/HHMI%20Interdisciplinary%20Collaboration%20for%20Bio%20Education%20Reform) to design a course in physics specifically for biology majors. NEXUS places the physics course second, after an introductory biology course. You can view, and even participate in, this work-in-progress via various comment pages (e.g., http://umdberg.pbworks.com/w/page/49797017/Readings%20Physics%20131; scroll down to the Overview section, where you will find a nice page of physics for biologists websites at http://umdberg.pbworks.com/w/page/35656531/Physics%20for%20Biologists%20Websites).

There are many biophysics courses and departments around the world. Wikiversity has a large comprehensive biophysics course online (http://en.wikiversity.org/wiki/Biophysics#Languages:_28en_2C_28fr_2C_28de_29) with some excellent graphics; text in French, English, Spanish, and German; and some sections available in numerous other languages. Phil Nelson at the University of Pennsylvania is the author of an undergraduate biophysics textbook and a leading proponent of meaningful curricular connections between biology and physics (www.physics.upenn.edu/~pcn). You can download a version of a presentation he calls “Keeping the Physics in Biophysics” (www.physics.upenn.edu/~pcn/1009Oxford.pdf); it is entertaining, informative, and provocative (see the quote I used in the opening paragraph of this Feature). Nelson’s presentation explores what he thinks are some of the best opportunities to present interesting and important biophysics to undergraduates, primarily in the areas of molecular biophysics, genetics, and bioinformatics, but also in neural processing and imaging. As a finale, you can become inspired by nature after reading Mary Salimi’s excellent review article on designs and structures that engineers have derived or copied from living organisms (www.scribd.com/doc/120688728/Biomimetics-and-bioinspiration-for-design-of-innovative-materials-and-systems).

ACKNOWLEDGMENTS

Thanks to Malcolm Campbell for helpful editorial comments.

REFERENCE

Biologists have long been concerned about the quality of undergraduate biology education. Indeed, some biology education journals, such as the American Biology Teacher, have been in existence since the 1930s. Early contributors to these journals addressed broad questions about science learning, such as whether collaborative or individual learning was more effective and the value of conceptualization over memorization. Over time, however, biology faculty members have begun to study increasingly sophisticated questions about teaching and learning in the discipline. These scholars, often called biology education researchers, are part of a growing field of inquiry called discipline-based education research (DBER).

DBER investigates both fundamental and applied aspects of teaching and learning in a given discipline; our emphasis here is on several science disciplines and engineering. The distinguishing feature of DBER is deep disciplinary knowledge of what constitutes expertise and expert-like understanding in a discipline. This knowledge has the potential to guide research focused on the most important concepts in a discipline and offers a framework for interpreting findings about students’ learning and understanding in that discipline. While DBER investigates teaching and learning in a given discipline, it is informed by and complementary to general research on human learning and cognition and can build on findings from K–12 science education research.

DBER is emerging as a field of inquiry from programs of research that have developed somewhat independently in various disciplines in the sciences and engineering. Although biology education research (BER) has emerged more recently than similar efforts in physics, chemistry, or engineering education research, it is making contributions to the understanding of how students learn and gain expertise in biology. These contributions, together with those that DBER has made in physics and astronomy, chemistry, engineering, and the geosciences, are the focus of a 2012 report by the National Research Council (NRC, 2012)\(^1\). For biologists who are interested in education research, the report is a

---

1To download a free PDF version of the report, visit www.nap.edu/catalog.php?record_id=13362.
useful reference, because it offers the first comprehensive synthesis of the emerging body of BER and highlights the ways in which BER findings are similar to those in other disciplines.

In this essay, we draw on the NRC report to highlight some of the insights that BER in general and BER in particular have provided into effective instructional practices and undergraduate learning, and to point to some directions for the future. The views in this essay are ours as editors of the report and do not represent the official views of the Committee on the Status, Contributions, and Future Directions of Discipline-Based Education Research; the NRC; or the National Science Foundation (NSF).

CHALLENGES TO UNDERGRADUATE LEARNING IN SCIENCE AND ENGINEERING

DBER and related research on teaching and learning have illuminated several challenges undergraduate students face in learning science and engineering. Indeed, “these challenges can pose serious barriers to learning and acquiring expertise in a discipline, and they have significant implications for instruction, especially if instructors are not aware of them” (NRC, 2012, p. 191).

One major challenge is accurate conceptual understanding. In every discipline, students have incorrect ideas and beliefs about concepts fundamental to the discipline. They particularly struggle with the unseen and with very small or very large spatial and temporal scales, such as those involved in understanding the interaction of subatomic particles or natural selection. As an example, many students believe the mass of a tree trunk comes from the soil, rather than the CO2 in the air, because they have difficulty believing that air has mass (Koba and Tweed, 2009).

Students’ incorrect knowledge poses a challenge to learning, because it comes in many forms, ranging from a single idea to a flawed mental model that is based on incorrect understandings of several interrelated concepts (Chi, 2008). It is less complicated to identify and address incorrect understandings of single ideas (e.g., all blood vessels have valves) than flawed mental models (e.g., the human circulatory system is a single loop rather than a double loop). Still, given that our goal is to help students progress toward more expert-like understandings, it is important for instructors to be aware of the misunderstandings that stand in the way of that goal and to have strategies for addressing those misunderstandings.

Understanding and using representations such as equations, graphs, models, simulations, and diagrams pose another major challenge for undergraduate students. Developing expertise in a discipline includes becoming familiar with representations unique to that discipline, such as evolutionary trees in biology, depictions of molecular structures in chemistry, and topographic maps in the geosciences. Experts in a discipline (here, professors) have long since mastered these representations and might no longer remember a time when these equations and images were new and confusing. However, in every discipline of science and engineering, students have difficulty understanding, interpreting, and creating representations that are unique and central to a given domain.

SOME INSTRUCTIONAL STRATEGIES FOR IMPROVING LEARNING AND CONCEPTUAL UNDERSTANDING

DBER has shown that specific instructional strategies can improve students’ learning and understanding. For example, the use of “bridging analogies” can help students bring incorrect beliefs more in line with accepted scientific explanations in physics (Brown and Clement, 1989). With bridging analogies, instructors provide a series of links between a student’s correct understanding and the situation about which he or she harbors an erroneous understanding. Another approach, interactive lecture demonstrations—in which students predict the result of a demonstration, discuss their predictions with their peers, watch the demonstration, and compare their predictions with the actual result—have been shown to improve students’ conceptual understanding in chemistry and physics (Sokoloff and Thornton, 1997).

DBER and related research also point to several strategies that can be used to improve students’ ability to use and understand diagrammatic representations. To this end, Hegarty (2011) suggests that instructors might:

- Explicitly point out the relationship among different displays of the same information to help students see the similarities.
- Explain the strengths and weaknesses of different representations for different purposes.
- Provide extensive opportunities for students to practice creating and interpreting diagrams of the desired type.

More generally, DBER and related research provide compelling evidence that student-centered instructional strategies can positively influence students’ learning, achievement and knowledge retention, as compared with traditional instructional methods, such as lecture. These strategies include asking questions during lecture and having students work in groups to solve problems, make predictions, and explain their thinking to one another. As noted in the NRC report on DBER, the point is not to abandon lecture entirely, but to use a range of carefully chosen instructional approaches that can include lecture. When lectures are used, they should be designed with attention to how best they can support students’ learning.

Despite compelling evidence for the effectiveness of student-centered approaches such as interactive lectures and collaborative activities, these practices still are not widespread among science and engineering faculty. In fact, science and engineering faculty are more likely than faculty in other disciplines to rely on lecture (Jaschik, 2012). Considering the many factors that influence decisions about instructional practices, it is not hard to understand why many faculty members hesitate to embrace more interactive classroom approaches. Even those who are interested in adopting research-based instructional methods might find challenges in departments and institutions that do not provide the needed supports for faculty to change their practices, from students who are resistant to change, and in reward systems that do not prioritize teaching. Still, with support from colleagues, professional societies, and others, many faculty members have overcome these and other challenges to transform their instructional practices.
THE CONTRIBUTIONS OF BER

What role has BER played in identifying students’ challenges in learning biology and in helping to promote the use of research-based practices among biology faculty members? Most BER since the mid-1990s has focused on identifying students’ conceptual understandings, developing concept inventories that measure students’ understanding of a given concept, and studying the effectiveness of different types of instructional approaches that promote greater student engagement (Dirks, 2011). BER scholars use a variety of methods to study these problems. Depending on the questions being examined, these methods range from interview studies or classroom observations with a few or perhaps dozens of students, to quantitative comparisons of learning gains made with different instructional approaches across many courses or institutions. Much of this research focuses on students in the first 2 years of their undergraduate careers, typically in classroom settings in the context of large, introductory courses—the setting that provides the greatest challenge for generating engagement.

As the examples in the preceding sections illustrate, research in BER has produced some important insights into learning and, in some cases, guidance for improving teaching. A notable case of the latter comes from evolutionary biology, a field in which cognitive scientist Laura Novick and biologist Kefyn Catley have conducted extensive research about how students understand evolutionary relationships when different types of evolutionary tree representations are used (Catley and Novick, 2008; Novick et al., 2010). Their research shows that the form of representation that is most commonly used in undergraduate biology texts leads to the least understanding of this important evolutionary concept. As a result of their research, almost all introductory biology texts have now been changed to more effectively support undergraduate learning of evolutionary relationships, impacting the learning of hundreds of thousands of students each year.

These contributions notwithstanding, many opportunities exist to enhance the value of BER, and of DBER more generally. For example, despite the importance of fieldwork to biology, comparatively little BER has been conducted in the field. Other emerging areas of research in DBER—and in BER by extension—include longitudinal studies, studies that examine similarities and differences among different student groups, research related to the affective domain and the transfer of learning, and the development of assessments to measure student learning. According to the NRC’s 2012 report on DBER, a specific challenge for BER scholars is to “identify instructional approaches that can help overcome the math phobia of many biology students and introduce more quantitative skills into the introductory curriculum, as computational biology and other mathematical approaches become more central to the field of biology” (NRC, 2003).

As BER grows, clarity about supporting BER scholars versus implementing BER findings to improve undergraduate biology education will be helpful. Regarding the support of BER scholars, the Society for the Advancement of Biology Education Research (SABER) provides a venue for BER scholars to share their research and support the development of early-career BER scholars. Several life sciences professional societies, including the American Society for Microbiology, the American Society for Neuroscience, already offer professional development opportunities for faculty members to consider how to integrate BER findings into their teaching; others could use these models to do the same.

Findings from BER studies are increasingly accessible to those who are interested in using them to inform their teaching, as well as to those who might be interested in pursuing BER research programs. BER scholars publish their research on teaching and learning in a wide variety of journals. In a review of the BER literature from 1990–2010, Clarissa Dirks (2011) identified ~200 empirical studies on college students’ learning, performance or attitudes. Although these articles appeared in more than 100 different journals, most were published in just four: the Journal of Research in Science Teaching, the Journal of College Science Teaching, Advances in Physiology Education, and CBE—Life Sciences Education (LSE). The past decade has seen a particularly rapid increase in the number of BER articles, especially in LSE.

Regarding the implementation of BER findings to improve undergraduate biology teaching, efforts are under way in several disciplines to help increase current and future faculty members’ use of research-based practices. In biology, two notable examples are the National Academies Summer Institute for Undergraduate Education in Biology and the NSF-sponsored Faculty Institutes for Reforming Science Teaching (FIRST) program. The Summer Institute works with teams of university faculty, emphasizing the application of teaching approaches based on education research, or “scientific teaching.” FIRST supports postdoctoral students interested in strengthening their teaching approaches. Although participants of the Summer Institute workshops reported substantial increases in their use of research-based instructional strategies over time (Pfund et al., 2009), an analysis of videotaped lessons from participants of the Summer Institute and the FIRST Program yielded mixed results concerning changes in practices (Ebert-May et al., 2011). It is important to note that alumni of the Summer Institute frequently reported that it took three or more years of experimentation before they could effectively implement learner-centered strategies (Pfund et al., 2009). As the NRC’s 2012 report concludes, “These results suggest that measuring the influence of DBER and related research on teaching requires a nuanced, longitudinal model of individual behavior rather than a traditional ‘cause and effect’ model using a workshop or other delivery mechanism as the intervention” (p. 173).

Individual scholars in the BER community can promote the acceptance and use of BER findings to improve undergraduate biology learning in two significant ways. One way is to enhance the quality of BER. As with any field, DBER has strengths and limitations. The greatest strength of DBER is the contribution of deep disciplinary knowledge to questions of teaching and learning in a discipline. In all disciplines, DBER could be enhanced by linking to other bodies of relevant research (including BER in other disciplines), being explicitly grounded in theories of teaching and learning, using standardized measures for assessing learning gains and student attitudes, and conducting research on a larger scale than a single classroom and over longer periods of time than a single course. To link to other bodies of research, BER scholars could ask their BER colleagues in physics, chemistry, and the geosciences to review draft manuscripts. SABER could
help by establishing mechanisms to connect BER scholars to DBER studies in other disciplines; examples exist in engineering and the geosciences. And journal editors and reviewers could encourage the authors of BER articles to include citations of similar work in related fields.

BER scholars also can help to promote change at the departmental and institutional levels without assuming responsibility for sweeping reforms. Relatively straightforward strategies include disseminating key findings to colleagues or getting together on campus to discuss and strategize possible changes. BER scholars seeking a more active role in promoting institutional change might also help department chairs understand how to evaluate the research of BER faculty.

Given the unusually large number of diverse life sciences professional societies, the emerging coherence and focus of the biology undergraduate community on BER and improving learning in biology is notable. The growing body of BER literature and the professionalization of the field in the context of SABER in less than half a decade are cause for celebration. The American Association for the Advancement of Science Vision and Change in Undergraduate Biology (http://visionandchange.org) efforts and the associated Vision and Change Leadership Fellows program (www.pulsecommunity.org) to drive department-level change in biology education emphasize implementation of widespread adoption of BER findings. The trajectory is promising.

REFERENCES


Feature
Book Review

Genesis of What Is Life?: A Paradigm Shift in Genetics History


Reviewed by John R. Jungck, Departments of Biological Sciences and Mathematical Sciences, University of Delaware, Newark, DE 19716

Most geneticists can easily name a dozen papers that transformed their field. Many of these papers continue to be reprinted and commented upon by historians, philosophers, literary critics, science technology and society scholars, and biologists. Such papers include Mendel’s 1866 plant hybridization paper; Sutton’s 1902 synthesis of Mendel with meiosis, which thereby coupled chromosomal mechanics with independent assortment; Morgan’s 1910 paper on sex linkage; Wright’s 1930 “Evolution in Mendelian Populations”; Luria and Delbrück’s 1942 paper on the fluctuation test; and, of course, Watson and Crick’s 1953 paper on the structure of DNA. A seminal piece that is often absent from these lists is What Is Life?: The Physical Aspect of the Living Cell, in which Erwin Schrödinger (1944) offers his interpretation of the “Three-Man Paper” (3MP). The three men who authored the 3MP in its original German, “Über die Natur der Genmutation und der Genstruktur,” were Nikolai Timofeeff-Ressovsky, Karl Zimmer, and Max Delbrück. While Delbrück, a theoretical physicist, went on to win a Nobel Prize for his later work with Salvador Luria, the other two authors were of equal stature in their fields of experimental genetics and experimental radiology. Schrödinger’s book, rather than the original 3MP, has caught the attention of many geneticists, who argue that it prompted physicists to move into biology after World War II and instigated the hunt for an “aperiodic crystal” and a “code script” that many felt led to the discovery that DNA was the hereditary molecule and the deciphering of the genetic code.

Creating a Physical Biology is a collection by a series of authors who return to the source—situating the 3MP in its own historical period. Well-prefaced and edited by Sloan and Fogel, the essays in this collection highlight the differences...
between the actual 3MP and Schrödinger’s text. They relate the 3MP to issues of teleology and reductionism, especially whether physicists like Bohr really expected biology to have fundamental new physical laws. They also summarize mechanistic conceptions of life, various misreadings of the paper, and the development of biophysics. In my view, the treatment of these topics reinforces the importance of interdisciplinarity, theory construction, and careful mathematical analysis, as well as the value of collaboration among workers from different educational backgrounds.

The collection includes Fogel’s extraordinarily readable English translation of the 3MP (pp. 221–271). I urge readers to read this first to get a sense of the prescience of this seminal paper and to then read the “Translator’s Preface” (pp. 214–220). This will avoid a “prefiltering” (Sankaran, 2012) by the commentators and translator and afford the treat of seeing Timofe´eff-Ressovsky’s clever experiments to determine the impact of x-rays in inducing mutations by making use of Nobel laureate Hermann J. Muller’s “C1B” and “attached X” crossing methods for identifying mutants in Drosophila melanogaster. Timofe´eff-Ressovsky’s experiments investigated mutation rates as a function of x-ray dosage, wavelength, and temperature. He also distinguished spontaneous from induced mutations, concluding that induced mutations only affected one of the two alleles and that some mutations were reversible. In the second section of the paper, Zimmer articulates the famous “target theory,” which posited a set of simplifying assumptions that accounted for the probabilistic interaction between radiation and the target (genes). Delbrück then synthesizes much of the preceding experimental work and uses analogies from quantum mechanical thinking to make interpretations about the nature of genes. He concludes that genes are macromolecules with a specific atomic composition (“a well-defined assemblage of atoms”) that can be altered by x-radiation and states, “We want to emphasize that the fundamental property of the gene [is] its identical self-replication during mitosis . . . .” Finally, all three coauthor the final section of the article to draw their general conclusions: “The genome is a highly complicated physical-chemical structure, consisting of a series of specific, chemical pieces of matter—the individual genes. . . . it leads to an explicit or implicit critique of the cell theory; the cell, thus far from proving itself so magnificently as the unit of life, dissolves into the ‘ultimate units of life,’ the genes.” Wow! I had no idea that such language and specificity preceded Schrödinger. I am embarrassed that I never investigated the antecedent before.

Michael A. Goldman (2011) says: “Sloan and Fogel argue that Schrödinger’s What Is Life? misrepresents the 3MP. They note, for instance, that he misleads by saying that quantum mechanics makes possible ‘a complete reduction of biological to physical systems,’ which the paper never claims. Schrödinger also ignores its reservations about mapping genotype to phenotype. But there is little evidence that he intended to provide an authentic account.” Because Schrödinger’s What Is Life? is so responsible for most geneticists’ understanding of the 3MP, such re-examination of its claims is important.

An important addition to this wonderful collection is the work of Alexander von Schwerin (2010), who gives a detailed reappreciation of the target theory. Furthermore, he helps us appreciate the transition in our thinking over the past century: He puts the 3MP into further context by tracing the subsequent history of approaches to understanding mutagenesis. In particular, he notes that Charlotte Auerbach, in 1969, substantially changed our view of mutagenesis from being an aberrant chemical or physical (x-ray) damage to being a normal biological process. “It is a task of its own to draw that historic line of ‘physiologization’ of mutations—and, hence, of the activation of the organism as an actor of its own in the process of the transformation of external stimuli into mutations.” With Miroslav Radman’s work on trade-offs between accuracy and efficacy, which led to a resolution of the neo-Lamarckians’ attack on Darwinism with their adaptive mutation hypothesis by showing demonstrably that the rate of production of mutations due to mismatch repair systems or fidelity of DNA replication is related to the harshness and/or stability of a population’s environment, this important distinction becomes all the more important.

Was the 3MP crucial to a paradigm shift? The wonderful translation and commentary provided by Creating a Physical Biology has convinced me that it did. Furthermore, while the “target theory” turned out to be incorrect, the synthetic research program laid out a heuristic approach to future practice and set an extraordinarily high standard for the synthesis of experiment, theory, and mathematical hypothesis testing. Furthermore, Creating a Physical Biology offers insight into a moment in time when such a substantial shift in thinking occurred and speculation on why and how it did.

REFERENCES


The Other Half of the Story: Effect Size Analysis in Quantitative Research

Jessica Middlemis Maher, Jonathan C. Markey, and Diane Ebert-May

Department of Plant Biology, Michigan State University, East Lansing, MI 48824-1312

INTRODUCTION

Quantitative research in biology education is primarily focused on describing relationships between variables. Authors often rely heavily on analyses that determine whether the observed effect is real or attributable to chance, that is, the statistical significance, without fully considering the strength of the relationship between those variables (Osbourne, 2008). While most researchers would agree that determining the practical significance of their results is important, statistical significance testing alone may not provide all information about the magnitude of the effect or whether the relationship between variables is meaningful (Vaske, 2002; Nakagawa and Cuthill, 2007; Ferguson, 2009).

In education research, statistical significance testing has received valid criticisms, primarily because the numerical outcome of the test is often promoted while the equally important issue of practical significance is ignored (Fan, 2001; Kotrlik and Williams, 2003). As a consequence, complete reliance on statistical significance testing limits understanding and applicability of research findings in education practice.

Effect Size and Statistical Significance Testing: Why Both Are Necessary

Imagine that a researcher set up two treatment conditions: for example, unfertilized and fertilized plants in a greenhouse or, similarly, reformed and traditional teaching approaches in different sections of an introductory biology course. The

Therefore, authors and referees are increasingly calling for the use of statistical tools that supplement traditionally performed tests for statistical significance (e.g., Thompson, 1996; Wilkinson and American Psychological Association [APA] Task Force on Statistical Inference, 1999). One such tool is the confidence interval, which provides an estimate of the magnitude of the effect and quantifies the uncertainty around this estimate. A similarly useful statistical tool is the effect size, which measures the strength of a treatment response or relationship between variables. By quantifying the magnitude of the difference between groups or the relationship among variables, effect size provides a scale-free measure that reflects the practical meaningfulness of the difference or the relationship among variables (Coe, 2002; Hojat and Xu, 2004).

In this essay, we explain the utility of including effect size in quantitative analyses in educational research and provide details about effect size metrics that pair well with the most common statistical significance tests. It is important to note that effect size and statistical significance testing (which we will shorten to “significance testing,” also known as hypothesis testing) are complementary analyses, and both should be considered when evaluating quantitative research findings (Fan, 2001). To illustrate this point, we begin with two hypothetical examples: one in biology and one in education.

CBE—Life Sciences Education
Vol. 12, 345–351, Fall 2013
A researcher is interested in knowing whether the first treatment is more or less effective than the second, using some measurable outcome (e.g., dried plant biomass or student performance on an exam); this constitutes the research hypothesis. The null hypothesis states that there is no difference between the treatments. Owing to sampling variation in a finite sample size, even if the two treatments are equally effective (i.e., the null hypothesis is true), one sample mean will nearly always be greater than the other. Therefore, the researcher must employ a statistical significance test to determine the probability of a difference between the sample means occurring by chance when the null hypothesis is true. Using the appropriate test, the researcher may determine that sampling variability is not a likely explanation for the observed difference and may reject the null hypothesis in favor of the alternative research hypothesis. The ability to make this determination is afforded by the statistical power, which is the probability of detecting a treatment effect when one exists, of the significance test. Statistical power is primarily determined by the size of the effect and the size of the sample: as either or both increase, the significance test is said to have greater statistical power to reject the null hypothesis.

The basis for rejection of the null hypothesis is provided by the p value, which is the output of statistical significance testing that is upheld as nearly sacred by many quantitative researchers. The p value represents the probability of the observed data (or more extreme data) given that the null hypothesis is true: Pr(occurred by chance | H0), assuming that the sampling was random and done without error (Kirk, 1996; Johnson, 1999). A low value of p, typically below 0.05, usually leads researchers to reject the null hypothesis. However, as critics of significance testing have pointed out, the abuse of this rather arbitrary cutoff point tends to reduce the decision to a reject/do not reject dichotomy (Kirk, 1996). In addition, many researchers believe that the smaller the value of p, the larger the treatment effect (Nickerson, 2000), equating the outcome of significance testing to the importance of the findings (Thompson, 1993). This misunderstanding is likely due to the fact that when sample size is held constant, the value of p correlates with effect size for some statistical significance tests. However, that relationship completely breaks down when sample size changes. As described earlier, the ability of any significance test to detect a fixed effect depends entirely on the statistical power afforded by the size of the sample. Thus, for a set difference between two populations, simply increasing sample size may allow for easier rejection of the null hypothesis. Therefore, given enough observations to afford sufficient statistical power, any small difference between groups can be shown to be “significant” using a statistical significance test.

The sensitivity of significance testing to sample size is an important reason why many researchers advocate reporting effect sizes and confidence intervals alongside test statistics and p values (Kirk, 1996; Thompson, 1996; Fan, 2001). Kotrlik and Williams (2003) highlight a particularly clear example in which statistical and practical significance differ. In their study, Williams (2003) was interested in comparing the percent time that faculty members spend teaching with the percent time that they would prefer to spend teaching. Despite the fact that the mean differences between actual and preferred teaching time were statistically significant ($t_{154} = 2.20$, $p = 0.03$), the effect size (Cohen's $d = 0.09$) was extremely small (see Tables 1 and 2 for effect size metrics and interpretations). As a result, the author did not suggest that there were practically important differences between actual and preferred teaching time commitments (Williams, 2003). Reporting the confidence interval would have also illustrated

<table>
<thead>
<tr>
<th>Table 1. Common measures of effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Effect size measure</strong></td>
</tr>
<tr>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Odds ratio</td>
</tr>
<tr>
<td>Cohen’s $d$</td>
</tr>
<tr>
<td>Hedges’ $g$</td>
</tr>
<tr>
<td>Glass’s $\Delta$</td>
</tr>
<tr>
<td>Cohen’s $f$</td>
</tr>
<tr>
<td>Eta-squared</td>
</tr>
<tr>
<td>Partial eta-squared</td>
</tr>
<tr>
<td>Pearson’s $r$</td>
</tr>
<tr>
<td>Point-biserial correlation coefficient ($r_{pb}$)</td>
</tr>
</tbody>
</table>
the small effect in this study: while the confidence interval would not have contained zero, one of its end points would have been very close to zero, suggesting that the population mean difference could be quite small.

Although Williams (2003) presents a case in which a small “significant” p value could have led to an erroneous conclusion of practically meaningful difference, the converse also occurs. For example, Thomas and Janues (1996) present an example from a study of juvenile rainbow trout willingness to forage under the risk of predation (Johnsson, 1993). An important part of the study tested the null hypothesis that large and small juveniles do not differ in their susceptibility to the predator, an adult trout. Using eight replicate survivorship trials, Johnsson (1993) found no significant difference in the distribution of risk between the two size classes (Wilcoxon signed-rank test: $T^+ = 29, p = 0.15$). However, the data suggest that there may in fact be a biologically significant effect: on average, 19 ± 4.9% (mean ± SE) of the large fish and 45 ± 7% of the small fish were killed by the predator (Johnsson, 1993). This difference likely represents a medium effect size (see Table 2; Thomas and Janues, 1996). Not reporting effect size resulted in the researchers failing to reject the null hypothesis, possibly due to low statistical power (small sample size), and the potential to erroneously conclude that there were no differences in relative predation risk between size classes of juvenile trout.

Thus, metrics of effect size and statistical significance provide complementary information: the effect size indicates the magnitude of the observed effect or relationship between variables, whereas the significance test indicates the likelihood that the effect or relationship is due to chance. Therefore, interpretations derived from statistical significance testing alone have the potential to be flawed, and inclusion of effect size reporting is essential to inform researchers about whether their findings are practically meaningful or important. Despite the fact that effect size metrics have been available since the 1960s (Huberty, 2002) and have been recognized as being a potentially useful aspect of analyses since the 1990s (e.g., Cohen, 1994; Thompson, 1996; Wilkinson and APA Task Force on Statistical Inference, 1999), the adoption of effect size as a complement to significance testing has been a slow process, even in high-impact research (Tressoldi et al., 2013). Nevertheless, many journals are beginning to develop editorial policies requiring some measure of effect size to be reported in quantitative studies (e.g., Royer, 2000). In response to this need for implementation, we next discuss the various methods used to calculate effect sizes and provide guidance regarding the interpretation of effect size indices.

<table>
<thead>
<tr>
<th>Table 2. Interpreting effect size valuesa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect size measure</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Odds ratio</td>
</tr>
<tr>
<td>Cohen’s $d$ (or one of its variants)</td>
</tr>
<tr>
<td>$r$</td>
</tr>
<tr>
<td>Cohen’s $f$</td>
</tr>
<tr>
<td>Eta-squared</td>
</tr>
</tbody>
</table>


Measures of Effect Size: Two Categories

We concentrate on parametric tests and group effect sizes into two main categories: those for 1) comparing two or more groups and 2) determining strength of associations between variables. The most frequently used statistical tests in these two categories are associated with specific effect size indices (see Table 1; Cohen, 1992), and we will discuss some of the more common methods used for each below. Refer to Figure 1 for a general guide to selecting the appropriate effect size measure for your data.

Comparing Two or More Groups. A common approach to both biological and educational research questions is to compare two or more groups, such as in our earlier examples comparing the effects of a treatment on plant growth or student performance. For these kinds of analyses, the appropriate measure of effect size will depend on the type of data collected and the type of statistical test used. We present here a sample of effect size metrics relevant to $\chi^2$, t, or F tests.

When comparing the distribution of a dichotomous variable between two groups, for instance, when using a $\chi^2$ test of homogeneity, the odds ratio is a useful effect size measure that describes the likelihood of an outcome occurring in the treatment group compared with the likelihood of the outcome occurring in the control group (see Table 1; Cohen, 1994; Thompson, 1996). An odds ratio equal to 1 means that the odds of the outcome occurring is the same in the control and treatment groups. An odds ratio of 2 indicates that the outcome is two times more likely to occur in the treatment group when compared with the control group. Likewise, an odds ratio of 0.5 indicates that the outcome is two times less likely to occur in the treatment group when compared with the control group. Granger et al. (2012) provide an example of reporting odds ratios in educational research. In their study, the effectiveness of a new student-centered curriculum and aligned teacher professional development was compared with a control group. One of the instruments used to measure student outcomes produced dichotomous data, and the odds ratio provided a means for reporting the treatment’s effect size on this student outcome. However, the odds ratio alone does not quantify treatment effect, as the magnitude of the effect depends not only on the odds ratio but also on the underlying value of one of the odds in the ratio. For example, if a new treatment for an advanced cancer increases the odds of survival by 50% compared with the existing treatment, then the odds ratio of survival is 1.5. However, if oddscontrol = 0.002 and oddstreatment = 0.003, the increase is most likely not practically meaningful. On the other hand, if an oddscontrol = 0.5 and oddstreatment = 0.75, this could be interpreted as a substantial increase that one might find practically meaningful.

When comparing means of continuous variables between two groups using a t test, Cohen’s $d$ is a useful effect size measure that describes the difference between the means normalized to the pooled standard deviation (SD) of the two groups (see Table 1; Cohen, 1988). This measure can be used only when the SDs of two populations represented by the two groups are the same, and the population distributions are close to normal. If the sample sizes between the two groups differ significantly, Hedges’ $g$ is a variation of Cohen’s $d$ that can be used to weight the pooled SD based on sample sizes (see Table 1 for calculation; Hedges, 1981). If the SDs of the populations differ, then pooling the sample SDs is not
appropriate, and other ways to normalize the mean difference should be used. Glass’s $\Delta$ normalizes the difference between two means to the SD of the control sample (see Table 1). This method assumes that the control group’s SD is most similar to the population SD, because no treatment is applied (Glass et al., 1981). There are many relevant examples in the educational research literature that employ variations on Cohen’s $d$ to report effect sizes. Abraham et al. (2012) used Cohen’s $d$ to show how an instructional treatment affected students’ post scores on a test of the acceptance of evolutionary theory. Similarly, Matthews et al. (2010) used Cohen’s $d$ to show the magnitude of change in student’s beliefs about the role of mathematics in biology due to changes in course materials, delivery, and assessment between different years of the same course. Gottesman and Hoskins (2013) applied Cohen’s $d$ to compare pre/post means of data collected using an instrument measuring students’ critical thinking, experimental design ability, attitudes, and beliefs.

When comparing means of three or more groups, for instance, when using an analysis of variance (ANOVA) test, Cohen’s $f$ is an appropriate effect size measure to report (Cohen, 1988). In this method, the sum of the deviations of the sample means from the combined sample mean is normalized to the combined sample SD (see Table 1). Note that this test does not distinguish which means differ, but rather just determines whether all means are the same. Other effect size measures commonly reported with ANOVA, multivariate analysis of covariance (MANCOVA), and analysis of covariance (ANCOVA) results are eta-squared and partial eta-squared. Eta-squared is calculated as the ratio of the between-groups sum of squares to the total sum of squares (see Table 1; Kerlinger, 1964). Alternatively, partial eta-squared is calculated as the ratio of the between-groups sum of squares to the sum of the between-groups sum of squares and the error sum of squares (Cohen, 1973). For example, Quitadamo and Kurtz (2007) reported partial eta-squared, along with ANCOVA/MANCOVA results, to show effect sizes of a writing treatment on student critical thinking. However, eta-squared is deemed by some as a better measure to report, because it describes the variance accounted for by the dependent measure (Levine and Hullett, 2002), which bears similarities to typical measures reported in correlational studies.

**Determining Strength of Association between Variables.** Another common approach in both biological and educational research is to measure the strength of association between two or more variables, such as determining the factors that predict student performance on an exam. Many researchers using this type of analysis already report appropriate measures of effect size, perhaps without even realizing they are doing so. In most cases, the regression coefficient or analogous index provides information regarding the magnitude of the effect.

The Pearson product-moment correlation coefficient (Pearson’s $r$) measures the association between two continuous variables, such as in a linear regression (see Table 1). Squaring the $r$ value when performing a simple linear regression results in the coefficient of determination ($r^2$), a measure that provides information about the amount of variance shared between the two variables. For multiple-regression analysis, the coefficient of multiple determination ($R^2$) is an appropriate effect size metric to report. If one of the study variables is dichotomous, for example, male versus female or pass versus fail, then the point-biserial correlation coefficient ($r_{pb}$) is the appropriate metric of effect size. The point-biserial correlation coefficient is similar in nature to Pearson’s $r$ (see Table 1). An easy-to-use Web-based calculator to calculate $r_{pb}$ is located at www.vassarstats.net/pbcorr.html. Spearman’s rank
correlation coefficient (\( \rho \)) is a nonparametric association measure that can be used when both variables are measured on an ordinal or ranked scale or when variables on a continuous scale are not normally distributed. This measure can be used only after one applies a transformation to the data that ranks the values. Because this is a nonparametric measure, Spearman’s \( \rho \) is not as sensitive to outliers as Pearson’s \( r \). Note that there are also variations of Spearman’s \( \rho \) that handle different formats of data. Most statistical software packages can calculate all of these measures of variable association, as well as most of the measures comparing differences between groups. However, one must be careful to be sure that values provided by the software are indeed what they are claimed to be (Levine and Hullett, 2002).

**How to Interpret Effect Sizes**

Once you have calculated the effect size measure, how do you interpret the results? With Cohen’s \( d \) and its variants, mean differences are normalized to SD units. This indicates that a \( d \) value of 0.5 can be interpreted as the group means differing by 0.5 SDs. Measures of association report the strength of the relationship between the independent and dependent variables. Additional manipulation of these association values, for example, \( r^2 \), can tell us the amount of shared variance between the variables. For the case of regression analysis, we can assume that an \( r^2 \) value of 0.3 means that 30% of the variance in the dependent variable can be explained by the independent variable. Additionally, McGraw and Wong (1992) developed a measure to report what they call “the common language effect size indicator,” which describes the probability that a random value sampled from one group will be greater than a random value sampled from a comparison group (McGraw and Wong, 1992).

Statisticians have determined qualitative descriptors for specific values of each type of effect size measure (Cohen, 1988, 1992; Rosenthal, 1996). For more interpretation of these types of measures, see Table 2. These values can help guide a researcher to make some sort of statement about the qualitative nature of the effect size, which is useful for communicating the meaning of results. Additionally, effect size interpretations impact the use of data in meta-analyses. Please refer to Box 1 to see an example of how interpretations of the different types of effect size measures can be converted from one type to another for the purpose of meta-analysis.

**Limitations of Effect Size**

We have built a justification for the reporting of effect sizes as a complement to standard statistical significance testing. However, we do not wish to mislead the reader to construe effect size as a panacea in quantitative analyses. Effect size indices should be used and interpreted just as judiciously as \( p \) values. Effect sizes are abstract statistics that experience biases from sampling effort and quality and do not differentiate among relationships of similar magnitude that may

### Box 1. Use of effect sizes in meta-analyses

Effect size measures are an important tool used when performing meta-analyses because they provide a standardized method for comparing results across different studies with similar designs. Two of the more common measures are Pearson’s \( r \) and Cohen’s \( d \). Cohen’s \( d \) describes the difference between the means of two groups normalized to the pooled standard deviation of the two groups. Pearson’s \( r \) measures the association between two continuous variables. A problem arises when comparing a study that reports an \( r \) value with one that reports a \( d \) value. To address this problem, statisticians have developed methods to convert \( r \) values into \( d \) values, and vice-versa. The equations are listed below:

\[
d = \frac{r}{\sqrt{1-r^2}}, \quad r = \frac{d}{\sqrt{d^2+4}}
\]

Many studies in the literature do not report effect sizes, and only report statistical significance results such as \( p \) values. Rosenthal and Rubin (2003) have developed a measure to account for this issue, \( r_{\text{equivalent}} \), which can determine effect size from experimental designs comparing the means of two groups on a normally distributed outcome variable (Rosenthal and Rubin, 2003). This measure allows meta-analysis researchers to derive apparent effect sizes from studies that only report \( p \) values and sample sizes. First, one determines a \( t \) value from a \( t \)-value table by using the associated sample size and one-tailed \( p \) value. Using this \( t \) value, one can calculate \( r_{\text{equivalent}} \) using the following equation:

\[
r_{\text{equivalent}} = \sqrt{\frac{t^2}{t^2+df}}, \quad \text{where } df = \text{degrees of freedom on which the } p\text{-value is based.}
\]

### Table 3. Recommended references for learning more about and implementing effect size measures as a part of standard statistical analyses

| Introduction to effect sizes written for the nonstatistician and relevant to the educational researcher |
| Theoretical explanation of effect size measures written for those with stronger statistical foundation |
| Accessible and relevant reference for the practical application of effect size in quantitative research; includes directions for calculating effect size in SPSS |
| A guide to implementing effect size analyses written for the researcher |
| American Psychological Association recommendation to report effect size analyses alongside statistical significance testing |
actually have more or less practical significance (Coe, 2002; Nakagawa and Cuthill, 2007; Ferguson, 2009). Rather, determination of what constitutes an effect of practical significance depends on the context of the research and the judgment of the researcher, and the values listed in Table 2 represent somewhat arbitrary cutoffs that are subject to interpretation. Just as researchers may have logical reasons to choose an alpha level other than $p = 0.05$ with which to interpret statistically significant, the interpretation of practical relationships based on effect size may be more or less conservative, depending on the context. For example, an $r$ of 0.1 for a treatment improving survival of a fatal disease may be of large practical significance. Furthermore, as we mentioned earlier, one should always accompany the proper effect size measure with an appropriate confidence interval whenever possible (Cohen, 1994; Nakagawa and Cuthill, 2007; Ellis, 2010; Tressoldi et al., 2013). For example, Lauer et al. (2013) reported Cohen’s $d$ along with 95% confidence intervals to describe the effects of an administration of a values-affirmation exercise on achievement gaps between men and women in introductory science courses.

CONCLUSION

By highlighting the problems with relying on statistical significance testing alone to interpret quantitative research results, we hope to have convinced the reader that significance testing is, as Fan (2001) puts it, only one-half of the coin. Our intent is to emphasize that no single statistic is sufficient for describing the strength of relationships among variables or evaluating the practical significance of quantitative findings. Therefore, measures of effect size, including confidence interval reporting, should be used thoughtfully and in concert with significance testing to interpret findings. Already common in such fields as medical and psychological research due to the real-world ramifications of the findings, the inclusion of effect size reporting in results sections is similarly important in educational literature. The measures of effect size described here do not by any means represent the numerous possible indices, but rather are intended to provide an overview of some of the most common and applicable analyses for educational research and a starting point for their inclusion in the reporting of results. In addition to the references cited throughout this article, we recommend several informative and accessible authorities on the subject of effect sizes, summarized in Table 3.

ACKNOWLEDGMENTS

We thank Alla Sikorskii for helpful comments and edits on an earlier draft of this essay.

REFERENCES


Many institutions require candidates for faculty positions to present a teaching demonstration as part of the interview process. To help job candidates prepare for this and to assist departments in planning how to structure this portion of the interview, we surveyed biology faculty from community and liberal arts colleges and master’s and PhD-granting institutions concerning how their departments assess teaching potential. We also asked survey respondents to share advice on how candidates can prepare for teaching demonstrations. Here we report on the survey results and offer suggestions based on comments from respondents.

INTRODUCTION

It is an all-too-familiar scenario: The job candidate for a biology department faculty position gives an outstanding research seminar, showing skill in formulating a hypothesis, carrying out meaningful research, analyzing data, forming conclusions, and translating work into the larger picture of science and society—it is an effective demonstration of the process of science. However, during the teaching demonstration portion of the interview, rather than engaging the audience in the learning process, the candidate delivers a presentation with too many slides, each of which is packed with factual information. The candidate interprets all the graphs and data for the audience, presents conclusions, and only intermittently asks whether the audience has any questions. Furthermore, the job candidate expects the audience to sit passively and absorb knowledge. In short, the candidate shows little ability to help others learn the process of science.

The wide discrepancy between the quality of the job candidate’s research talk and teaching demonstration indicates inadequate preparation for the teaching component of the job interview. There are a number of reasons that can explain lack of preparation, but two major ones are: 1) not knowing what is expected during the teaching demonstration and 2) a lack of effective practice of the desired skill. Though each job candidate has been attending class since he or she was a child, few have had enough practice teaching, and even fewer have delineated the crucial aspects of effective teaching.

To help job candidates better understand faculty expectations of the teaching demonstration and to help departments think about how to structure this portion of the interview, we canvassed 113 biology faculty from a variety of institutions across North America (Table 1) as to the role and assessment of the teaching demonstration in the interview process. We asked faculty who vote on tenure-track hiring decisions and are in departments in which a teaching demonstration is part of the interview process to identify the elements of an effective teaching demonstration and to give advice as to how candidates can prepare for this aspect of the interview. The results of the survey, as well as representative comments from survey respondents, are presented below.
Job Interview Teaching Demonstration

Table 1. Demographic information on survey respondents

<table>
<thead>
<tr>
<th>Institution type</th>
<th>Number of respondents</th>
<th>Current position</th>
</tr>
</thead>
</table>
| Community college             | 30                    | Full professor: 40%  
|                               |                       | Assistant professor: 13%  
|                               |                       | Lecturer: 37%  
|                               |                       | Other: 0%  |
| Primarily undergraduate institute | 35                | Full professor: 26%  
|                               |                       | Associate professor: 37%  
|                               |                       | Assistant professor: 31%  
|                               |                       | Lecturer: 3%  
|                               |                       | Other: 0%  |
| Master's degree granting      | 7                     | Full professor: 14%  
|                               |                       | Associate professor: 72%  
|                               |                       | Assistant professor: 14%  
|                               |                       | Lecturer: 0%  
|                               |                       | Other: 0%  |
| PhD granting                  | 41                    | Full professor: 35%  
|                               |                       | Associate professor: 26%  
|                               |                       | Assistant professor: 21%  
|                               |                       | Lecturer: 16%  
|                               |                       | Other: 2%  |

Table 2. The requirement for job candidates to do a teaching demonstration varies by institution type

<table>
<thead>
<tr>
<th>Institution type</th>
<th>just a research talk (%)</th>
<th>a research talk and a teaching demonstration (%)</th>
<th>just a teaching demonstration (%)</th>
<th>Other (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community college</td>
<td>0</td>
<td>6.7</td>
<td>83.3</td>
<td>10.0</td>
</tr>
<tr>
<td>Primarily undergraduate institute</td>
<td>34.3</td>
<td>34.3</td>
<td>0</td>
<td>11.4</td>
</tr>
<tr>
<td>Master's degree granting</td>
<td>42.9</td>
<td>57.1</td>
<td>0</td>
<td>19.5</td>
</tr>
<tr>
<td>PhD granting</td>
<td>43.9</td>
<td>36.6</td>
<td>0</td>
<td>15.1</td>
</tr>
</tbody>
</table>

Table 3. The importance of the teaching demonstration relative to the research talk at different institution types

<table>
<thead>
<tr>
<th>Institution type</th>
<th>more weight than his/her performance on the research talk (%)</th>
<th>less weight than his/her performance on the research talk (%)</th>
<th>equal weight with his/her performance on the research talk (%)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primarily undergraduate institute</td>
<td>41.7</td>
<td>0</td>
<td>58.3</td>
<td>12</td>
</tr>
<tr>
<td>Master's degree granting</td>
<td>50.0</td>
<td>25.0</td>
<td>25.0</td>
<td>4</td>
</tr>
<tr>
<td>PhD granting</td>
<td>0</td>
<td>53.3</td>
<td>46.7</td>
<td>15</td>
</tr>
</tbody>
</table>

Table 4. The importance of the teaching demonstration relative to the research talk at different institution types

<table>
<thead>
<tr>
<th>Institution type</th>
<th>more weight than his/her performance on the research talk (%)</th>
<th>less weight than his/her performance on the research talk (%)</th>
<th>equal weight with his/her performance on the research talk (%)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primarily undergraduate institute</td>
<td>41.7</td>
<td>0</td>
<td>58.3</td>
<td>12</td>
</tr>
<tr>
<td>Master's degree granting</td>
<td>50.0</td>
<td>25.0</td>
<td>25.0</td>
<td>4</td>
</tr>
<tr>
<td>PhD granting</td>
<td>0</td>
<td>53.3</td>
<td>46.7</td>
<td>15</td>
</tr>
</tbody>
</table>

Another reason faculty want to hire candidates who demonstrate potential as effective teachers is that, since 2000, a number of national reports have made calls to action to improve science teaching (National Research Council, 2000, 2003; American Association for the Advancement of Science, 2010; Anderson et al., 2011). These reports conclude that learning is most effective when it is an active endeavor incorporating inquiry-based learning strategies and integrating all steps of the scientific process into the learning process. A wealth of studies indicate that few students learn while sitting passively in lecture taking notes (e.g., Bonwell and Eison, 1991; Bransford et al., 2000; Knight and Wood, 2005; Ruiz-Primo et al., 2011). Therefore, faculty are being asked to recognize
that students must “do” science to “learn” science and their teaching methods should reflect this change.

Finally, showing teaching potential demonstrates not only that a candidate is prepared for the major task of teaching, but also that he or she is well prepared for establishing his or her own research projects. Just as research enhances teaching, it is also true that teaching enhances research. A recent study of science graduate students found that those who teach inquiry-based methods improve their research skills in formulating hypotheses and designing experiments to test these hypotheses (Feldon et al., 2011). Giving a skilled teaching demonstration therefore establishes a candidate’s potential as both an effective teacher and a research scientist.

**Elements of an Effective Teaching Demonstration**

To determine which elements of a teaching demonstration are most important, we asked survey participants to respond to 21 statements describing specific characteristics of a teaching session on a Likert scale from 1 = not important to 4 = very important (Table 5).

The top-rated statement from faculty across institution types was that the content of a candidate’s teaching demonstration be accurate (Q21), which indicates that, first and foremost, faculty feel it is important that the job candidate have expertise in his or her discipline. Disciplinary expertise includes factual knowledge, as well as a deep understanding of the conceptual frameworks that underlie and connect these facts. However, to be an effective instructor, the candidate should also demonstrate pedagogical content knowledge, that is, the ability to select, organize, and properly implement the appropriate teaching method to help students meet the challenges of mastering the material (Shulman, 1986). Statements aligned with the importance of pedagogical content knowledge were also ranked highly in our survey, including: the presentation is understandable to students (Q12), the material is organized effectively (Q18), and the candidate pitches the talk at the correct level for the intended audience (Q4).

Our survey showed few differences based on institutional type in how faculty ranked the importance of specific elements in a teaching demonstration. Only two statements showed a significant difference: whether “the candidate discusses how he/she would assess student learning on an exam or other future assignment” (Q5: Kruskal-Wallis $\chi^2 = 6.11, df = 2, p < 0.05$) and whether “the candidate explains why he/she is using certain teaching strategies” (Q19:

<table>
<thead>
<tr>
<th>Institution type</th>
<th>How much do teaching and teaching evaluations determine tenure and promotion in your department?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Teaching does not contribute, only research effort is important (%)</td>
</tr>
<tr>
<td>Community college</td>
<td>0</td>
</tr>
<tr>
<td>Primarily undergraduate institution</td>
<td>0</td>
</tr>
<tr>
<td>Master’s degree granting</td>
<td>0</td>
</tr>
<tr>
<td>PhD granting</td>
<td>4.8</td>
</tr>
</tbody>
</table>

**Table 5. The ranked importance of elements of a teaching demonstration**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Mean score$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q21. The candidate’s content information is accurate.</td>
<td>3.86</td>
</tr>
<tr>
<td>Q12. The candidate’s presentation would be understandable to students.</td>
<td>3.82</td>
</tr>
<tr>
<td>Q18. The candidate is able to organize material effectively.</td>
<td>3.73</td>
</tr>
<tr>
<td>Q4. The candidate pitches the talk at the correct level for the intended audience.</td>
<td>3.68</td>
</tr>
<tr>
<td>Q7. The candidate introduces topics in a way that connects to the audience (asking questions, emphasizing the relevance of the topic, etc.).</td>
<td>3.63</td>
</tr>
<tr>
<td>Q9. The candidate demonstrates his/her knowledge about the topic.</td>
<td>3.60</td>
</tr>
<tr>
<td>Q1. The candidate is enthusiastic.</td>
<td>3.56</td>
</tr>
<tr>
<td>Q6. The candidate appears confident in his/her ability to teach.</td>
<td>3.52</td>
</tr>
<tr>
<td>Q8. The candidate covers the appropriate amount of material for the given time and level of student.</td>
<td>3.40</td>
</tr>
<tr>
<td>Q11. The candidate speaks at a comfortable pace.</td>
<td>3.22</td>
</tr>
<tr>
<td>Q10. The candidate’s slides are easy to read.</td>
<td>3.20</td>
</tr>
<tr>
<td>Q14. The candidate asks if the audience has questions during the presentation.</td>
<td>3.02</td>
</tr>
<tr>
<td>Q3. When asked a question, the candidate facilitates a discussion, rather than just telling the answer.</td>
<td>2.66</td>
</tr>
<tr>
<td>Q13. The candidate provides a wrap-up at the end.</td>
<td>2.62</td>
</tr>
<tr>
<td>Q16. The candidate allows wait time (at least 3–5 s) for the audience to think about questions posed.</td>
<td>2.58</td>
</tr>
<tr>
<td>Q17. The candidate incorporates elements of active learning (e.g. discussion, small-group work, clicker questions).</td>
<td>2.52</td>
</tr>
<tr>
<td>Q15. The candidate gives a clear indication of his/her teaching philosophy.</td>
<td>2.48</td>
</tr>
<tr>
<td>Q5. The candidate discusses how he/she would assess student learning on an exam or other future assignment.</td>
<td>2.26</td>
</tr>
<tr>
<td>Q19. The candidate explains why he/she is using certain teaching strategies.</td>
<td>2.02</td>
</tr>
<tr>
<td>Q2. The candidate brings in materials, such as a printout of slides, that he/she would hand out in class.</td>
<td>1.98</td>
</tr>
</tbody>
</table>

$^a$1 = not important to 4 = very important.

$^b$Based on the opinion of 49 respondents who vote on tenure-track hiring decisions and are in a department in which a teaching demonstration is part of the interview process.
Job Interview Teaching Demonstration

Kruskal-Wallis $\chi^2 = 8.51$, df $= 2$, $p < 0.05$). Whether a candidate discusses assessment methods elicited a greater range of responses from faculty at primarily undergraduate and PhD-granting institutions when compared with community college faculty (Figure 1A). Whether a candidate discusses teaching strategy elicited a greater range of responses from faculty at PhD-granting institutions when compared with the other two institution types (Figure 1B).

Elements of an Outstanding Teaching Demonstration

A teaching demonstration that incorporates all the statements shown in Table 5 would certainly be impressive. As one survey responder noted: “Very few candidates are able to meet all the criteria. If a candidate does at least, say, two-thirds of those well, then s/he is going to be ranked quite highly in my mind.” Given that the likelihood of a job candidate excelling at all the statements in Table 5 is small, we also wanted to determine which elements were key to making a teaching demonstration outstanding, to enable candidates to prioritize their teaching efforts.

To learn what faculty consider the key elements of an outstanding teaching demonstration, we asked the following: “What distinguishes an adequate teaching demonstration from an outstanding teaching demonstration?” The elements most frequently cited included: enthusiasm, passion for a topic, and a relaxed and confident manner.

The outstanding candidate conveys their excitement for the topic to their audience. The outstanding candidate also conveys to their students the sense that “I can do this,” along with the expectations to be achieved.

I ask myself if I would like to enroll in a course taught by the candidate. If I am left excited about the experience and left inspired and I feel my students feel the same I would say it was an outstanding demonstration. If I could tolerate attending the candidate’s class it would be considered adequate.

The use of active-learning strategies was also cited by respondents.

If the individual can demonstrate something out of the ordinary, for example, a hands-on activity or an online simulation, the presentation will clearly stand out.

An outstanding candidate will give us some indication that he/she will readily do more than lecture (some type of student centered teaching) upon arriving on campus.

Preparing for the Teaching Demonstration

Implementing the elements considered important in a teaching demonstration requires planning and practice. We have compiled a number of suggestions that can help job candidates as they prepare for the teaching demonstration part of their job interviews.

Follow Instructions. If you are given instructions by the department on the type of audience you are to be teaching, follow the instructions. Regardless of the makeup of the audience, it is advised that you treat them as though they were students. Furthermore, if you were told your teaching demonstration is to be pitched to an introductory biology class, do not give the same talk you would give in a graduate-level journal club.

In my experience, the teaching demonstration can easily fall into a continuation of the research presentation, centering on the presenter’s area of expertise rather than using a more balanced approach to the big picture that an undergraduate or even graduate student might need to progress.

Do Some Research. Once either you or the department you are visiting has selected a topic for your teaching demonstration, find a course at your home institution similar to the one your teaching demonstration will focus on. Sitting in on this class, looking over the course textbook, and talking with teaching assistants or faculty associated with the course will give you a much better understanding of what current students know and how to make complex material accessible to them. It will also give you insight into current classroom
Cover the Appropriate Amount of Material—Less IS More. When your audience is filled with faculty members, it is difficult to remember that you are not being judged on the amount of knowledge you can convey, but on how able you are to “teach” the process of science, using the subject you are discussing.

An outstanding demonstration would engage the audience (encouraging responses, discussion among the students), relate the material being taught to students’ interests and experience, [and] not view covering any given amount material as the main goal.

Although it is common for science teachers to feel the need to cover large amounts of content, effective teaching does not sacrifice depth, problem solving, and critical thinking in the process (Coil et al., 2010). Studies have shown there is a limit to the amount of information a person can process and store (Miller, 1956; Sweller, 1994). Realize, therefore, that your demonstration should only contain three to four major points that you present in a way that is accessible to students. Choose these well.

Engage Students in the Classroom. One key to facilitating learning is to engage students in the learning process (Bransford et al., 2000; deWinstanley and Bjork, 2002). We identify below key aspects of how to engage students in a manner that enhances learning.

Connect with the Audience: Help Them Realize the Importance and Relevance of the Topic. To capture the attention of your audience, relate the class material to something with which your audience is very familiar which provides some mystery or puzzle. Alluding to something in the recent or popular press is often quite effective in producing the “hook” that will encourage attention.

When considering material that could generate student interest, keep in mind that your class will contain a diversity of individuals. This diversity can cover ethnic, socioeconomic, religious, political, and gender differences. You want your teaching demonstration to engage but not offend any sector of this broad audience.

Use Slides and the Board to Promote Learning. Many teaching demonstrations include slides, such as PowerPoint slides, but keep in mind:

- PowerPoint is a good start for some, but it isn’t the whole talk and it isn’t essential. What is essential is that they get their point across.
- If you use slides in your teaching demonstration, each slide should help promote learning and display material so it is accessible to students. Your slides, therefore, should not be used as simply a way to convey information. For example, bulleted points are a classic way of reminding a speaker of what they want to say, but this is not an effective way to help students learn material. In fact, work in cognitive science has shown that supplying students with an extensive set of class notes actually impedes rather than enhances learning (deWinstanley and Bjork, 2002). Slides, instead, can be used to pose questions, query interpretations of graphs, and illustrate points with pictures and videos.

When incorporating graphs into your teaching demonstration, it is important that the graphs are large and readable and that all axes are labeled. Take time to orient the audience to the graph by asking the audience to explain what is depicted on each axis and to propose how the data were generated. Similarly, ask the audience to construct possible conclusions to be drawn from the data rather than just telling them your conclusions—interpretation of material is a key aspect of the student learning process (deWinstanley and Bjork, 2002).

If audience members should be writing down material you are presenting, it is best if you are writing it down as well. If a blackboard or whiteboard is available, use this to write down the information that you absolutely want the audience to write down.

It is important that the candidate uses the board, or other technique that slows the pace for students and allows for more spontaneity than only following (a) PPT.

Use Questions to Promote Learning: “Ask, Don’t Tell.” Using questions to introduce an idea is an effective way to focus the audience’s minds on the material that is coming. Your questions can be rhetorical or directed to elicit student discussion. In addition, it is important to verify that your audience is processing the information you are trying to convey. Job candidates will often stop periodically and ask whether there are any questions, which in a real classroom can be problematic, because students who are lost are usually reluctant to speak in front of the class. A more effective strategy is to ask a question that will diagnose whether the students actually understand the material. During your job interview, you could 1) pose a question, 2) have the audience members write down an answer and discuss their answers with their neighbors, and 3) ask the audience to share ideas. When responding to an audience member’s idea, it is a good idea to try to use at least a portion of each answer to build the correct answer, while also politely correcting any errors.

Many job candidates feel uncomfortable asking questions and fielding audience responses, because they are afraid they might not be able to quickly process an answer or handle follow-up questions. If a question is asked that is difficult to answer, give yourself time to thoughtfully consider the question by turning the question back to the audience. Ask the audience to “talk to your neighbor.” This approach will give you time to gather audience input to which you can add your own ideas. You do not have to know all the answers, but always acknowledge the value of a question, and tell the audience to research this out of class and bring their subsequent ideas to the next class.

Use Active-Learning Activities. Incorporating active-learning techniques into the classroom greatly enhances student learning (e.g., Bonwell and Eison, 1991; Bransford et al., 2000; Knight and Wood, 2005; Ruiz-Primo et al., 2011). These techniques encompass a wide variety of activities, such as small-group break-out discussions, group assignments, short in-class writing assignments, and use of personal response systems, such as clickers (Table 6 has a list of resources with information on active-learning strategies). Effectively incorporating any of these active-learning activities into your teaching demonstration can be impressive to the search committee.
Table 6. Resources to help job candidates become familiar with active learning

<table>
<thead>
<tr>
<th>Topic</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom assessment techniques</td>
<td>50 CATS by Angelo and Cross: Techniques for Assessing Course-Related Knowledge &amp; Skills: <a href="http://pages.uoregon.edu/tep/resources/newteach/fifty_cats.pdf">http://pages.uoregon.edu/tep/resources/newteach/fifty_cats.pdf</a></td>
</tr>
<tr>
<td>Clickers</td>
<td>Videos and instructor guide: <a href="http://www.cwsei.ubc.ca/resources/SEI_video.html">www.cwsei.ubc.ca/resources/SEI_video.html</a> Information on research relating to clickers: <a href="http://derekbruff.org/?page_id=2">http://derekbruff.org/?page_id=2</a></td>
</tr>
<tr>
<td>Workshops</td>
<td>FIRST IV for postdoctoral fellows: <a href="http://www.msu.edu/~first4/index.html">www.msu.edu/~first4/index.html</a> National Academies Summer Institute on Undergraduate Biology Education: <a href="http://www.academiessummerinstitute.org">www.academiessummerinstitute.org</a></td>
</tr>
<tr>
<td>Meetings</td>
<td>Society for the Advancement of Biology Education Research: <a href="http://saber-biologyeducationresearch.wikispaces.com">http://saber-biologyeducationresearch.wikispaces.com</a> Lilly Conference on College and University Teaching: <a href="http://www.iats.com/conferences/lilly-conference">www.iats.com/conferences/lilly-conference</a></td>
</tr>
</tbody>
</table>

Outstanding teaching demonstrations are distinguished by effective use of active learning exercises, ability to accommodate a variety of learning styles, and, most importantly, ability to engage the class.

If you mention active learning as an important part of your teaching philosophy statement, it is important that you use active-learning in your teaching demonstration.

One time, a candidate had an impressive statement of teaching philosophy that talked about employing state-of-the-art pedagogies, and then failed to use any active learning in the teaching demonstration. I found this particularly frustrating because it suggested an inability to recognize what good teaching/learning is.

If you are using electronic devices, such as clickers, however, make sure you are well versed in how they work before you use them for a job interview. If at all possible, take the opportunity to try them out in the room you will use for your teaching demonstration. Failing technology can doom a teaching demonstration, so either confirm all parts of the system are in working order or go with a technique that does not require technology, such as having audience members raise their hands.

If one tries to use methods or tools they are uncomfortable or unaccustomed to using merely to try to impress the committee, it will show, and work against them.

Practice! Just as you practice your research talk in front of peers, present your teaching demonstration to colleagues, especially lecturers and other faculty whose primary focus is on teaching. Seasoned teachers have a wealth of experience and information that can ensure that your teaching session is organized in a way that is accessible to students, accurate, and effective at helping students learn. If possible, also practice your talk in front of students at your home institution. Students are quite honest and direct about what they think makes effective teaching and can give your teaching demonstration a true test run.

When you give a practice talk, fully try out any active-learning strategies.

This mistake is avoidable. Practicing active-learning strategies will give you an idea of the range of responses you can expect from the audience and help you plan how to react. It is also important to set aside enough time for each active-learning activity. Job candidates often cut short their learning activity with the missive “in a real class I would give the students more time but I am cutting it short today due to time.” Be cautious about saying this, because such a statement may indicate poor time-management or an inability to implement the activity.

CONCLUSIONS

Above all, realize that departments want to hire someone who has the potential to be successful in the classroom. While success in research will be measured by your number of grants and publications, success in the classroom will be measured by how well you engage students in meaningful learning.

The outstanding demonstration of teaching helps us understand how the individual connects with students.

So, among all the things you are doing to prepare for your future career, capitalize on your teaching assignments and mentorship opportunities in your laboratory. Each of these endeavors offers the opportunity to learn how to interact effectively with students. Understanding student concerns and how to help students learn could be the key to landing your academic job.

ACKNOWLEDGMENTS

M.S. is supported in part by the National Science Foundation under grant 0962805. We thank the faculty who responded to this survey. Approval to evaluate faculty responses to survey questions (exempt status, protocol no. 2012-04-07) was granted by the institutional review board at the University of Maine.

REFERENCES


Miller GA (1956). The magic number seven plus or minus two: some limits on our capacity to process information. Psychol Rev 63, 81–97.


Developing a Test of Scientific Literacy Skills (TOSLS): Measuring Undergraduates’ Evaluation of Scientific Information and Arguments

Cara Gormally,*† Peggy Brickman,‡ and Mary Lutz§

*Georgia Institute of Technology, School of Biology, Atlanta, GA 30322; †Department of Plant Biology and §Department of Educational Psychology and Instructional Technology, University of Georgia, Athens, GA 30602

Submitted March 14, 2012; Revised July 19, 2012; Accepted July 19, 2012

Monitoring Editor: Elisa Stone

Life sciences faculty agree that developing scientific literacy is an integral part of undergraduate education and report that they teach these skills. However, few measures of scientific literacy are available to assess students’ proficiency in using scientific literacy skills to solve scenarios in and beyond the undergraduate biology classroom. In this paper, we describe the development, validation, and testing of the Test of Scientific Literacy Skills (TOSLS) in five general education biology classes at three undergraduate institutions. The test measures skills related to major aspects of scientific literacy: recognizing and analyzing the use of methods of inquiry that lead to scientific knowledge and the ability to organize, analyze, and interpret quantitative data and scientific information. Measures of validity included correspondence between items and scientific literacy goals of the National Research Council and Project 2061, findings from a survey of biology faculty, expert biology educator reviews, student interviews, and statistical analyses. Classroom testing contexts varied both in terms of student demographics and pedagogical approaches. We propose that biology instructors can use the TOSLS to evaluate their students’ proficiencies in using scientific literacy skills and to document the impacts of curricular reform on students’ scientific literacy.

INTRODUCTION

Science educators, scientists, and policy makers agree that development of students’ scientific literacy is an important aim of science education. Scientific literacy has been defined in multiple ways, all of which emphasize students’ abilities to make use of scientific knowledge in real-world situations (American Association for the Advancement of Science [AAAS], 1990, 2010; Bybee, 1993; Maienschein et al., 1998; Millar et al., 1998; DeBoer, 2000). For example, the National Research Council (NRC) defines scientific literacy as the ability “use evidence and data to evaluate the quality of science information and arguments put forth by scientists and in the media” (NRC, 1996). Project 2061 (AAAS, 1993) and the Programme for International Student Assessment describe scientific literacy as “the capacity to use scientific knowledge to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity” (Organisation for Economic Co-operation and Development, 2003). These two definitions are the framework for our working concept of scientific literacy.

Individuals use scientific information in many real-world situations beyond the classroom, in ways ranging from evaluating sources of evidence used in media reports about science to recognizing the role and value of science in society to interpreting quantitative information and performing...
quantitative tasks (Cook, 1977; Jenkins, 1990; Uno and Bybee, 1994; Koballa et al., 1997; Ryder, 2001; Kutner et al., 2007). Achieving scientific literacy for all is a core rationale for science coursework as part of general education (Gen Ed) requirements for undergraduates (Meinwald and Hildebrandt, 2010). In response to calls for reform and alignment with science education standards, instructors of these Gen Ed science courses have focused increasingly on students’ development of scientific literacy skills, including quantitative literacy (Quitadamo et al., 2008; Chevalier et al., 2010; Marsteller et al., 2010; Colon-Berlinger and Borrowes, 2011; Brickman et al., 2012). Coincident with this shift is an interest in finding ways to assess students’ development of scientific literacy skills, especially in the context of Gen Ed courses (Labov, 2004; DeHaan, 2005).

To date, several biology concept inventories have been developed to assess students’ conceptual knowledge (Anderson et al., 2002; Garvin-Doxas and Klymkowsky, 2008; Smith et al., 2008; Shi et al., 2010; Tsui and Treagust, 2010). However, similar progress has lagged in the realm of evaluating students’ scientific literacy skills as defined by the NRC standards (NRC, 1996). Researchers have yet to agree upon a single set of measurable skills critical for scientific literacy, beyond unanimously agreeing that these skills must include conceptual understanding, as well as views about science and society (Bauer et al., 2007). In a recent study surveying more than 150 life sciences faculty from a variety of institutions, faculty identified problem solving/critical thinking, oral and written communication, and the ability to interpret data as the three most important skills students should develop before graduation (Coil et al., 2010). However, these skills require further articulation into measurable constructs in order to effectively evaluate students’ mastery of these skills.

Several instruments have been developed to assess individual aspects of scientific literacy skills, but no single instrument measures all skills. Two surveys frequently used for international comparisons of scientific literacy include questions about non–lab-based science process skills, such as defining science, and items measuring vocabulary and basic content knowledge (Lemke et al., 2004; Milker, 2007). General scientific reasoning instruments were developed specifically to test cognitive skills related to critical thinking and reasoning (Lawson, 1978; Facione, 1991; Sundre, 2003, 2008; Sundre et al., 2008; Quitadamo et al., 2008). However, for the average instructor, too much time and money may be required to utilize multiple instruments to measure all these skills. Situational factors, such as large-enrollment courses, may also hamper the utility of these instruments. For example, an instrument recently developed by White et al. (2011) challenges subjects to confront issues of quality, credibility, and interpretation of scientific research using open-ended responses to conclusions from individual research studies, but this instrument may be challenging to use in large-enrollment courses. The lack of a readily accessible instrument for assessing students’ proficiency in evaluating scientific arguments and sources of evidence as described by the NRC may serve as a barrier to evaluating curricular reform (NRC, 1996).

Like other faculty who emphasize scientific literacy skills in classroom instruction, we had no ready means for evaluating the impact of a curricular reform in a large Gen Ed course. Several recent studies of students’ science literacy skill development in reformed biology courses have relied on multiple-choice exam questions (Fencl, 2010) or a pre- and postintervention test (Chevalier et al., 2010) as a means of assessment. In both cases, the test’s psychometric properties were unknown. Relying on tests such as these for evaluation purposes presents limitations for generalizing findings. To avoid this, we sought to develop a practical and psychometrically sound test for use across undergraduate introductory science courses. This test was designed to be freely available, as well as quick to administer and score.

We describe here the development process of the Test of Scientific Literacy Skills (TOSLS), as well as results from its use in Gen Ed biology courses at several different institutions. The instrument consists of 28 multiple-choice questions that are contextualized around real-world problems, for example, evaluating the reliability of an Internet site containing scientific information or determining what would constitute evidence to support a fitness product’s effectiveness. The TOSLS development process included articulating the skills critical for scientific literacy, examining the validity of the instrument through student interviews and biology educator expert reviews, pilot testing, subsequent examination of psychometric properties, and finally, classroom testing of the finalized instrument in multiple, different biology courses (Table 1).

<table>
<thead>
<tr>
<th>Table 1. Overview of TOSLS development process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Examined literature on existing instruments to identify scientific literacy skills</td>
</tr>
<tr>
<td>2. Conducted faculty survey to articulate what encompasses scientific literacy skills</td>
</tr>
<tr>
<td>3. Developed and administered a pilot assessment based on defined skills</td>
</tr>
<tr>
<td>4. Revised assessment based on item analyses and feedback from student interviews</td>
</tr>
<tr>
<td>5. Examined instrument validity through additional student interviews and biology faculty evaluations</td>
</tr>
<tr>
<td>6. Evaluated finalized instrument for item difficulty, item discrimination, and reliability</td>
</tr>
<tr>
<td>7. Administered instrument in multiple contexts to demonstrate utility and measured learning gains</td>
</tr>
</tbody>
</table>

INSTRUMENT DEVELOPMENT

Overview of TOSLS Development Process: Instrument Validity

Development of the TOSLS was an iterative process informed by the development process of recent instruments, including the Introductory Molecular and Cell Biology Assessment (Shi et al., 2010), Quantitative Reasoning Test and Scientific Reasoning Test (Sundre, 2003, 2008; Sundre et al., 2008), and CLASS Biology (Semsar et al., 2011). Establishing instrument validity was an important part of the development process. Validity determines the extent to which the instrument measures what it purports to measure (American Educational Research Association, 1999). We used multiple means to determine instrument validity, focusing on measures of content validity and construct validity (Osterlind, 2010). Content validity is the extent to which the instrument measures all facets of a given social construct, in this case, skills essential for scientific literacy. Measures of content validity included building on national reports and a faculty survey about skills...
essential for scientific literacy and utilizing expert biology faculty evaluations. Construct validity involves statistical analyses to evaluate item validity and relationships between instrument items. Measures of construct validity included test item analyses, internal test consistency, and expert faculty biology evaluations of instrument items.

Content Validity

Insuring Inclusion of Major Facets of Scientific Literacy Skills. We began by identifying key definitions provided in education policy documents and reviews in order to define the major facets of scientific literacy for this instrument (AAAS, 1993, 2010; National Academy of Sciences, 1997; Organisation for Economic Co-operation and Development, 2003; Sundre, 2003; Picone et al., 2007; Holbrook and Ran-nikmae, 2009; Bray Speth et al., 2010). We also heeded recent reports that recommend incorporating quantitative concepts into undergraduate introductory science courses, since quantitative literacy provides a common language across scientific disciplines (NRC, 2003; Bialek and Botstein, 2004; Gross, 2004; Kutner et al., 2007; Karsai and Kampis, 2010). Students need to develop a broad set of skills to approach scientific phenomena quantitatively (NRC, 2003), as well as to apply basic quantitative concepts in their daily lives (Kutner et al., 2007). Quantitative literacy, as defined for adults’ daily lives by the National Assessment of Adult Literacy, is the “knowledge and skills required to perform quantitative tasks (i.e., to identify and perform computations, either alone or sequentially, using numbers embedded in printed materials),” which may include calculating a percentage to figure a tip at a restaurant or the amount of interest on a loan (Kutner et al., 2007). Using this literature review, we identified skills related to two major categories of scientific literacy skills: 1) skills related to recognizing and analyzing the use of methods of inquiry that lead to scientific knowledge, and 2) skills related to organizing, analyzing, and interpreting quantitative data and scientific information. We articulated the skills as measurable outcomes, herein referred to as TOSLS skills (Table 2).

Faculty Survey. Because expert agreement provides strong support for content validity, we sought to verify the consistency of the skills we articulated through our literature review with the opinions of faculty teaching Gen Ed courses. Alignment between these two sources would support the claim that we included major facets of scientific literacy, and, in addition, would provide evidence of utility for faculty beyond our own courses. To determine the degree of consistency, we designed an online survey to elicit feedback from faculty teaching Gen Ed biology courses nationwide (included in the Supplemental Material). Specifically, we asked faculty to list the three most important skills for scientific literacy and to rate the importance of the skills required for students to be considered scientifically literate (described in Table 2). Finally, we asked these faculty whether they currently teach and assess these skills. We sent this survey to life science faculty and postdocs, using email listservs from professional organizations (National Association of Biology Teachers, Association of Biology Laboratory Educators, and Society for the Advancement of Biology Education Research, among others) and textbook publishers (John Wiley & Sons and McGraw-Hill). Survey respondents (n = 188) hailed from throughout the United States and represented a wide variety of higher education institutions, with 34% from private colleges and universities, 20% from public 2-yr colleges, 20% from public state universities, 17% from public research universities, 5% from public state colleges, and 4% from public regional universities. The majority of faculty respondents (78%) teach at least some students who are nonscience majors. Of these faculty, 40% teach Gen Ed biology courses composed solely of nonscience majors, while 38% teach courses composed of both science and nonscience majors. The remaining faculty participants teach a mix of science majors, including biology majors (12%), and courses for biology majors only (10%).

All three coauthors individually read and classified the survey responses into categories. Through discussion, we clarified and consolidated the categories we identified. Finally, one coauthor (M.L.) classified each response into the agreed-upon categories; all three coauthors discussed uncertainties as they arose in the classification process. The three most important skills that faculty listed for Gen Ed biology students to demonstrate scientific literacy strongly corresponded to our TOSLS skills. Of all skills cited by faculty respondents, the most frequent responses were related to understanding the nature of science (NOS; 15.44%), with responses such as “understand what serves as evidence in science” and “differentiate between science and non-science,” which align with skill 1: identifying a valid scientific argument (Table 2). Similarly, faculty identified skills related to other aspects of NOS, with the second, third, and fourth most frequent responses closely corresponding with skill 4: understand elements of research design and how they impact scientific findings/conclusions (15.09%); skill 2: evaluate the validity of sources (13.21%); and skill 3: evaluate the use and misuse of scientific information (8.58%), respectively. Although there has been an emphasis recently on the importance of quantitative literacy, only 12.87% of all responses aligned with quantitative and graphing skills (skills 5, 6, 7, 8, 9). Responses categorized as specific content knowledge accounted for more responses than any other skill described (21.1%). Respondents were asked to identify the importance, on a scale from 1 (unimportant) to 5 (very important) for undergraduates in Gen Ed biology courses to develop each of the nine TOSLS skills, as well as whether they currently taught and assessed the skills (Figure 1). When prompted with the quantitative skills, faculty rated the importance of teaching quantitative skills equal to that of NOS skills. The majority of faculty agreed that the TOSLS skills are important for scientific literacy. A large majority of faculty report that they currently teach all these skills (≥58.7% teach all skills, with the exception of skill 8: understanding and interpreting basic statistics, which only 44.9% of faculty report teaching). However, faculty report that they assess their students’ proficiencies in using these skills at lower rates than they report teaching these skills (≥57.5% assess most skills, with the exception of skill 8, which only 40.1% assess, and skill 3, which 40.1% assess, and skill 2, which 49.1% assess). (All skills are described in Table 2.)
Table 2. Categories of scientific literacy skills

<table>
<thead>
<tr>
<th>Questions</th>
<th>Explanation of skill</th>
<th>Examples of common student challenges and misconceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Understand methods of inquiry that lead to scientific knowledge</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Identify a valid scientific argument</td>
<td>1, 8, 11</td>
<td>Recognize what qualifies as scientific evidence and when scientific evidence supports a hypothesis</td>
</tr>
<tr>
<td>2. Evaluate the validity of sources</td>
<td>10, 12, 17, 22, 26</td>
<td>Distinguish between types of sources; identify bias, authority, and reliability</td>
</tr>
<tr>
<td>3. Evaluate the use and misuse of scientific information</td>
<td>5, 9, 27</td>
<td>Recognize a valid and ethical scientific course of action and identify appropriate use of science by government, industry, and media that is free of bias and economic, and political pressure to make societal decisions</td>
</tr>
<tr>
<td>4. Understand elements of research design and how they impact scientific findings/conclusions</td>
<td>4, 13, 14</td>
<td>Identify strengths and weaknesses in research design related to bias, sample size, randomization, and experimental control</td>
</tr>
<tr>
<td><strong>II. Organize, analyze, and interpret quantitative data and scientific information</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Create graphical representations of data</td>
<td>15</td>
<td>Identify the appropriate format for the graphical representation of data given particular type of data</td>
</tr>
<tr>
<td>6. Read and interpret graphical representations of data</td>
<td>2, 6, 7, 18</td>
<td>Interpret data presented graphically to make a conclusion about study findings</td>
</tr>
<tr>
<td>7. Solve problems using quantitative skills, including probability and statistics</td>
<td>16, 20, 23</td>
<td>Calculate probabilities, percentages, and frequencies to draw a conclusion</td>
</tr>
<tr>
<td>8. Understand and interpret basic statistics</td>
<td>3, 19, 24</td>
<td>Understand the need for statistics to quantify uncertainty in data</td>
</tr>
<tr>
<td>9. Justify inferences, predictions, and conclusions based on quantitative data</td>
<td>21, 25, 28</td>
<td>Interpret data and critique experimental designs to evaluate hypotheses and recognize flaws in arguments</td>
</tr>
</tbody>
</table>

**Construct Validity**

*Item Development Built from Student Challenges.* Assessment items are most valuable if they can assist in documenting students’ initial confusions, incomplete understandings, and alternative conceptions (Tanner and Allen, 2005). Therefore, we began our development of test items by reviewing studies that documented common student challenges in addressing problems relating to our set of scientific literacy...
skills (Table 2). We focused primarily on reviewing literature concerning postsecondary education. We would direct interested readers to review literature on student misconceptions at the K–12 level as well, since undergraduates may continue to hold these misconceptions. Many TOSLS skills involved recognizing and analyzing methods of inquiry that lead to scientific knowledge. Students must be able to critique scientific experiments, data, and results in order to make decisions about the ill-structured problems common to science. This, in its entirety, can be thought of as analyzing the strength of evidence-based arguments. We utilized the findings that students have trouble both formulating claims backed by evidence and providing reasoning for claims (Bray Speth et al., 2010), as well as linking claims to specific evidence (Cho and Jonassen, 2002) to begin construction of short-answer questions in these areas.

Critiquing the quality of sources of evidence is also an integral part of analyzing the strength of scientific arguments. The Internet has revolutionized access to scientific information for the average person and at the same time has exacerbated the need to critically evaluate these sources. In fact, 40% of U.S. Internet users report obtaining most of their scientific information from the Internet, and 87% of users report having searched online about science at least once (Horrigan, 2006). Students of all ages (primary, secondary, and higher education) encounter difficulties when evaluating the relevance and reliability of Web information (Ma Künstler et al., 2002; Brand-Gruwel et al., 2009). Left to their own devices, very few Internet users check the source and date of the information they find (Fox, 2006).

Credibility issues, such as recognizing conflicts of interest, affiliations, and expertise in sources of evidence, are also challenging for students. Even when introduced to Web evaluation criteria and asked to rank the quality of sites, students have difficulty evaluating sites for credibility and accuracy, instead using surface markers, such as currency, author, and amount and type of language used (Britt and Aglinskas, 2002; Walraven et al., 2009). Students often think the number of authors on a publication increases the credibility, thinking that each author adds independent corroboration of results (Brem et al., 2011). And students rarely venture beyond the initial site for independent corroboration, instead using surface markers, such as dates of posting and presence of details and percentages as evidence of accuracy (Brem et al., 2011). Students with low topic knowledge are more likely to trust poor sites and fail to differentiate between relevant and irrelevant criteria when judging the trustworthiness of sources (Braten et al., 2011). For this reason, we also included in our pilot assessment short-answer items that asked students to evaluate the quality of information from online sources, such as websites.

The TOSLS includes skills need to interpret numerical information (Table 2). This is also an integral part of acquiring functional scientific literacy, because scientific claims are often supported by quantitative data (Steen, 1997). Students have difficulty representing quantitative data on graphs, including labeling axes correctly and choosing the appropriate type of graph to display particular kinds of findings (Bray Speth et al., 2010). Students also have difficulty summarizing trends from data with variation, interpreting the biological meaning of a slope of a line, and interpreting graphs with interactions (Preece and Janvier, 1992; Bowen et al., 1999; Picone et al., 2007; Colon-Berlinger and Borrowes, 2011). For these reasons, we constructed multiple-choice items based on common student responses. For example, we adapted short-answer graphing questions used by Picone et al. (2007) into multiple-choice questions and provided students with short-answer questions that asked them to interpret information from graphs commonly seen in media reports found in periodicals such as the New York Times. We suggest that interested readers explore curricular resources at the National Institute of Mathematical and Biological Sciences (2012), as well as a recent report describing best practices for integrating mathematics in undergraduate biology (Marsteller et al., 2010).

**Pilot Item Testing.** At the beginning of the Summer 2010 semester, we piloted items probing the challenges described above with students in Concepts in Biology, a Gen Ed biology course at a large research university in the southeast (n = 80). We administered two isomorphic test forms, each containing nine short-answer questions and 14 multiple-choice questions contextualized around authentic real-world problems, such as evaluating the trustworthiness of information found from various sources (including a fictitious website) or interpreting data on meat consumption trends over the past 20 yr from a New York Times article. Following the test administration, we analyzed students’ written responses and constructed multiple-choice responses from frequently occurring answers to the nine short-answer questions. The practice of developing distracters, wrong answers that students frequently choose, from students’ own words is a well-established strength of concept inventory development (Sadler, 1998; D’Avanzo, 2008). We also recruited student volunteers for audiotaped cognitive interviews (n = 2). We conducted cognitive interviews across three iterations of the instrument-development process to aid in item refinement. Cognitive interviewing is a method used to elucidate whether respondents comprehend and respond to items in the way that researchers intend (Willis, 2005). We used this method to help identify unexpected problems in the wording of questions prior to expanding its use. Interviews were conducted by two graduate student research collaborators using

![Figure 1](image-url)
an interview protocol that included asking student interviewees to identify unknown terminology and confusing wording in each question, as well as think-alouds in which students were asked to give their reasoning for answering questions. The graduate student research collaborators transcribed the audiotaped interviews to summarize issues raised by interviewees. Two coauthors (P.B. and M.L.), along with the two graduate student research collaborators, listened to and discussed the audiotaped interviews. Responses to interviews were used to inform further test revision. At the end of the Summer 2010 semester, the revised test forms were administered, each with 23 multiple-choice questions. We analyzed test forms for item difficulty, reliability, and test equivalence. Unreliable items (defined as point biserial correlation scores below 0.15) were revised or removed.

During Fall 2010 and Spring 2011, we piloted the further revised multiple-choice assessments in two large-enrollment Gen Ed biology courses (Fall: Organismal Biology, \( n = 340 \); Spring: Concepts in Biology and Organismal Biology, \( n = 498 \) pre, \( n = 378 \) post), administering each form pre- and post-course. After each administration of the instrument, we examined the performance of each test question based on such indicators as item difficulty and item discrimination. We also examined the quality of the distractors for each item, looking for nondistracters (i.e., distracters that were chosen by five or fewer students) and poorly discriminating distractors. Well-written distractors should be selected more often by students with less knowledge in the domain of interest compared with those students selecting the correct answer. Conversely, poorly discriminating distracters are selected by a large number of high performers, and do not differentiate effectively among students with high and low scores on the overall test. These distracters may be poorly written or unintentionally challenging. In this study, distracters were considered to be poor discriminators when the overall test score mean for the students choosing the distracter was equal to or above the mean score for students choosing the correct answer. Poorly performing test questions were revised or removed prior to subsequent administrations.

Finally, in Summer 2011, we condensed the assessment to one form, with 30 multiple-choice questions, and administered the assessment only in the beginning of the semester (\( n = 70 \)). One coauthor (P.B.) conducted an audiotaped focus group interview with students (\( n = 5 \)), to determine their reasoning through each question. The interview was transcribed. Focus group findings were used to revise distracter choices in two major ways. First, we revised answer choices to represent true misconceptions rather than confusing wording. Second, we removed terminology such as “peer review” and “unbiased,” which students said clued them in to answer a question correctly without a deep understanding. In total, the assessment was piloted and revised through five semester-long cycles. Example items are shown in Table 3, and the complete test and the test with the answer key are included in the Supplemental Material.

**Expert Faculty Evaluation**

Expert evaluation was critical to ensure construct and content validity. We utilized several rounds of expert faculty evaluation of the items. During Fall 2010, five expert biology educators took both isomorphic test forms. In addition to answering the test questions, they provided comments on comprehension, relevance, and clarity. We used their comments to further revise items and to determine whether items should be removed from the instrument for subsequent rounds during the Fall 2010, Spring 2011, and Summer 2011 semesters. Once the instrument was in its final form, faculty experts in biology education, external to the project, were recruited by email to evaluate the assessment during Summer 2011 (\( n = 18 \)). Criteria for expertise included teaching introductory biology at the university level to Gen Ed students and participation in one of two national professional development programs: Faculty Institutes for Reforming Science Teaching or the National Academies Summer Institute on Undergraduate Education in Biology at the University of Wisconsin. Experts evaluated each question for scientific accuracy, commented on question understandability (Table 4), and answered each question themselves (Table 5). This set of evaluations guided final instrument development, serving in particular as a means to identify questions and answer items that required revision or removal and guiding the recategorization of items according to skills measured.

**Student Interviews**

During Fall 2011, student volunteers were solicited immediately following the pre- and postadministration of the instrument for think-aloud cognitive interviews (Willis, 2005). We selected students representing the diversity of the class, using information they provided about their major, gender, age, and experience in science courses. This included similar numbers of men and women from a variety of majors (education, humanities, business, math, and social sciences). A doctoral research assistant in math and science education experienced in interviewing techniques conducted individual hour-long, semi-structured interviews with 16 undergraduates (\( n = 10 \) at the beginning of the semester and \( n = 6 \) at the end of the semester). Each student volunteer was given a copy of the TOSLS and was asked to reflect and verbally articulate the reasoning process he or she used to answer each question. The interviews were audiotaped and transcribed by the graduate assistant. Two coauthors (C.G. and P.B.) followed a systematic approach to determine what characteristics would constitute correct reasoning for each skill set of questions, attempting to determine all components that would define correct reasoning. Together, they analyzed each student response, focusing on responses provided for correct answers to the multiple-choice questions. Any discrepancies were discussed until a consensus was reached. At this preliminary stage, three general types of student responses were identified: responses that provided correct reasoning, either describing why the student chose the correct multiple-choice answer and/or why the student excluded other answers; responses that were too vague to determine whether they provided correct reasoning; and responses indicating incorrect reasoning. Responses that were too vague (e.g., “It seems to be the only one”) were noted but excluded from further analysis. Using this rubric of three general student responses for each skill set of questions, the raters coded the full data set. Any student response that could not be classified according to correct reasoning as defined by the rubric was subject to
Test of Scientific Literacy Skills

Table 3. Example questions contextualized around real-world issues

Skill 1: Identifying a valid scientific argument
Question 1: Which of the following is a valid scientific argument?

a. Measurements of sea level on the Gulf Coast taken this year are lower than normal; the average monthly measurements were almost 0.1 cm lower than normal in some areas. These facts prove that sea level rise is not a problem.
b. A strain of mice was genetically engineered to lack a certain gene, and the mice were unable to reproduce. Introduction of the gene back into the mutant mice restored their ability to reproduce. These facts indicate that the gene is essential for mouse reproduction.
c. A poll revealed that 34% of Americans believe that dinosaurs and early humans coexisted because fossil footprints of each species were found in the same location. This widespread belief is appropriate evidence to support the claim that humans did not evolve from ape ancestors.
d. This winter, the northeastern United States received record amounts of snowfall, and the average monthly temperatures were more than 2°F lower than normal in some areas. These facts indicate that climate change is occurring.

Skill 2: Evaluate the validity of sources
Question #10: Your interest is piqued by a story about human pheromones on the news. A Google search leads you to the following website:

For this website which of the following characteristics is most important in your confidence that the resource is accurate or not.

a. The resource may not be accurate, because appropriate references are not provided.
b. The resource may not be accurate, because the purpose of the site is to advertise a product.
c. The resource is likely accurate, because appropriate references are provided.
d. The resource is likely accurate, because the website’s author is reputable.

Skill 3: Evaluate the use and misuse of scientific information
Question 9: Which of the following is not an example of an appropriate use of science?

a. A group of scientists who were asked to review grant proposals based on their funding recommendations on the researcher’s experience, project plans, and preliminary data from the research proposals submitted.
b. Scientists are selected to help conduct a government-sponsored research study on global climate change based on their political beliefs.
c. The Fish and Wildlife Service reviews its list of protected and endangered species in response to new research findings.
d. The Senate stops funding a widely used sex-education program after studies show limited effectiveness of the program.

Skill 4: Understand elements of research design and how they impact scientific findings/conclusions
Question 4: Which of the following research studies is least likely to contain a confounding factor (variable that provides an alternative explanation for results) in its design?

a. Researchers randomly assign participants to experimental and control groups. Females make up 35% of the experimental group and 75% of the control group.
b. To explore trends in the spiritual/religious beliefs of students attending U.S. universities, researchers survey a random selection of 500 freshmen at a small private university in the South.
c. To evaluate the effect of a new diet program, researchers compare weight loss between participants randomly assigned to treatment (diet) and control (no diet) groups, while controlling for average daily exercise and predict weight.
d. Researchers tested the effectiveness of a new tree fertilizer on 10,000 saplings. Saplings in the control group (no fertilizer) were tested in the fall, whereas the treatment group (fertilizer) were tested the following spring.

Skill 5: Create graphical representations of data
Question 15: Researchers found that chronically stressed individuals have significantly higher blood pressure compared with individuals with little stress. Which graph would be most appropriate for displaying the mean (average) blood pressure scores for high-stress and low-stress groups of people?

Skill 6: Read and interpret graphical representations of data
Question 18: Which of the following is the most accurate conclusion you can make from the data in this graph?

a. The largest increase in meat consumption has occurred in the past 20 yr.
b. Meat consumption has increased at a constant rate over the past 40 yr.
c. Meat consumption doubles in developing countries every 20 yr.
d. Meat consumption increases by 50% every 10 yr.
Table 3. Continued

Skill 7: Solve problems using quantitative skills, including probability and statistics

Question #23: A gene test shows promising results in providing early detection for colon cancer. However, 5% of all test results are falsely positive; that is, results indicate that cancer is present when the patient is, in fact, cancer-free. Given this false positive rate, how many people out of 10,000 would have a false positive result and be alarmed unnecessarily?

a. 5
b. 35
c. 50
d. 500

Skill 8: Understand and interpret basic statistics

Question 19: Two studies estimate the mean caffeine content of an energy drink. Each study uses the same test on a random sample of the energy drink. Study 1 uses 25 bottles, and study 2 uses 100 bottles. Which statement is true?

a. The estimate of the actual mean caffeine content from each study will be equally uncertain.
b. The uncertainty in the estimate of the actual mean caffeine content will be smaller in study 1 than in study 2.

c. The uncertainty in the estimate of the actual mean caffeine content will be larger in study 1 than in study 2.
d. None of the above

Table 4. Summary of expert responses to the three queries about the 28 TOSLS questions

<table>
<thead>
<tr>
<th>Subject of query</th>
<th>Agreement of experts (n = 18)</th>
<th>Number of questions (n = 28)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&gt;90%</td>
<td>&gt;80%</td>
</tr>
<tr>
<td>The information given in this question is scientifically accurate.</td>
<td>23</td>
<td>5</td>
</tr>
<tr>
<td>The question is written clearly and precisely.</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td>After taking a college Gen Ed science course, students should be able to answer</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>this question.</td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

Discussion about whether the list should be amended. Through this iterative process of synchronous rating and discussion, the rubric was refined (see the Supplemental Material). Finally, a single rater (C.G.) coded all the responses, using the rubric, and determined the frequencies of responses in all categories. In total, students answered 81.25% of the multiple-choice questions correctly. In terms of the reasoning students provided for the correctly answered multiple-choice questions, 4.4% of responses were vague and 94.5% of responses provided correct reasoning.

Table 5. Mean pre- and posttest scores of students from each course with calculated t value and effect size, as well as scores from biology faculty experts

<table>
<thead>
<tr>
<th>Course</th>
<th>Mean % correct (SE)</th>
<th>t</th>
<th>Effect size</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project-based nonmajors at public research university</td>
<td>61.71 (1.05)</td>
<td>10.51*</td>
<td>0.83</td>
<td>0.734</td>
<td>0.758</td>
</tr>
<tr>
<td>Traditional nonmajors at public research university</td>
<td>58.33 (0.99)</td>
<td>6.65*</td>
<td>0.48</td>
<td>0.718</td>
<td>0.713</td>
</tr>
<tr>
<td>Private research university</td>
<td>84.63 (1.30)</td>
<td>0.32</td>
<td>0.03</td>
<td>0.581</td>
<td>0.632</td>
</tr>
<tr>
<td>Midsized state college</td>
<td>44.29 (1.70)</td>
<td>1.22</td>
<td>0.12</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Biology majors at public research university</td>
<td>61.72 (0.71)</td>
<td>7.65*</td>
<td>0.33</td>
<td>0.682</td>
<td>0.761</td>
</tr>
<tr>
<td>Biology experts</td>
<td>N/A</td>
<td>91.43 (0.98)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*Pre- and posttest internal consistency is shown.

b*p < 0.05 (indicates significant gains).
Test of Scientific Literacy Skills

Figure 2. (a) Pre- and postmeasures of item difficulty, with results from the nonscience majors in the lecture-based section and (b) the project-based section of Concepts in Biology in Fall 2011 (* \( p < 0.05 \) difference between pre- and posttest scores). Questions are grouped according to skills (Table 2).

Statistical Characterization

After multiple rounds of pilot testing, individual and focus group interviews of student think-alouds, and expert reviews, we administered the final version of the TOSLS to students taking Concepts in Biology, an introductory biology class for nonmajors taught using a traditional lecture-based format (referred to herein as the “traditional nonmajors” course; \( n = 296 \)). The psychometric properties of pre- and postsemester administrations of the TOSLS included item difficulty (Figure 2), item discrimination (Figure 3), and test reliability (Crocker and Algina, 2008; Osterlind, 2010).

Item difficulty measures the proportion of the total sample that answered a question correctly. Item difficulties range from 0 to 1.0, with larger values representing “easier” test items. Individual item difficulties ranging from 0.30 to 0.80 are acceptable, particularly when difficulties are symmetrically distributed across a test (Feldt, 1993). The average item difficulty for the TOSLS was 0.59 on the pretest and 0.68 on the posttest (Figure 2). Item difficulties ranged from 0.32 to 0.88 on the pretest and 0.33 to 0.91 on the posttest.

Item discrimination indices quantify how well a test question differentiates among students with high and low scores on the overall test. Students with well-developed scientific literacy skills, for example, should be more likely to answer test items correctly than students with poorly developed skills. Item discrimination scores for the TOSLS were calculated using corrected point biserial correlations. Item discrimination scores below 0.20 indicate that the item poorly differentiates among students with high and low abilities (Ebel, 1965). The average item discrimination for the TOSLS was between 0.26 and 0.27 for the pre- and posttests, respectively (Figure 3). Item discrimination indices ranged from 0.05 to 0.36 on the pretest and from 0.09 to 0.41 on the posttest.

The overall reliability of the TOSLS was explored by examining the internal consistency of the test. Internal consistency estimates indicate the degree to which a group of items measure the same construct. We used the Kuder-Richardson 20 formula, a measure of internal consistency appropriate for use with binary data. Internal consistency estimates above 0.70 are considered acceptable, and values above 0.8 are considered to reflect good test reliability (Cronbach, 1951). The internal reliability of the TOSLS was 0.731 and 0.748 on the pretest and posttest, respectively (Table 5). These scores fall within the acceptable range of reliability. An exploratory factor analysis, a principal components analysis with a Varimax rotation, indicated that one factor rather than two or more

Figure 3. Pre- and postmeasures of item discrimination from Fall 2011. Findings from lecture-based and projects-based sections are shown combined.
factors best accounted for the variance in the data. These results indicate that the tested skills are related and that it is meaningful to view a student’s score on the TOSLS as a measure of his or her scientific literacy skills.

Instrument Administration and Measurement of Learning Gains

During Fall 2011, the multiple-choice question assessment was administered pre- and postsemester at three types of undergraduate institutions: a public research university, a private research university, and a midsized state college (Table 6). We chose to administer the instrument at several different institutions, with pedagogical approaches ranging from primarily lecture-based to reformed learner-centered courses. In administering the TOSLS, we sought to demonstrate the test’s utility across multiple contexts, as well as to determine the sensitivity of the TOSLS to highlight differences in learning gains. The assessment was administered in two different courses at a large public research university (very high research activity), with a primarily residential student body, located in the southeastern United States. One section of Concepts of Biology, a Gen Ed biology course, was taught primarily through lecturing (traditional nonmajors course), while the other section of the course was taught using a project-based applied-learning (PAL) curriculum (project-based nonmajors course; described in Brickman et al., 2012). The assessment was administered in a second course at public research university, Principles of Biology I, an introductory lecture-based course for biology majors (referred to herein as “biology majors” course). The assessment was administered in Introduction to Environmental Biology, a Gen Ed biology course taught using PAL, required for environmental biology majors, but only an elective credit for biology majors, at a private large research university (very high research activity), with a highly residential student body, located in the midwest. Finally, the assessment was administered in Principles of Biology, a primarily lecture-based Gen Ed biology course at a midsized state college, a masters-granting medium-sized public college, located in the southeast, with a primarily non-residential student body. Institutions were chosen by convenience sampling through word-of-mouth at research conferences and responses to our faculty survey. Consequently, we were able to implement the TOSLS in Gen Ed courses at different types of institutions and to work with faculty interested and committed to using the TOSLS for one semester in their courses.

Mean differences in pre- and posttest scores were examined using paired-sample t tests for each class. Effect sizes, which quantify the magnitude of mean differences in standardized terms, were also calculated (Cohen’s $d = t \cdot [2(1 − r)/n]^{1/2}$) (Dunlap et al., 1996; Andrews et al., 2011; Table 5). Results indicated that posttest scores were significantly higher than pretest scores for the three classes (i.e., project-based, traditional, and biology majors) at the public research university, according to results from paired-sample t tests. Examination of effect sizes revealed that learning gains were large in magnitude for the project-based nonmajors class, approaching medium in magnitude for the traditional nonmajors class, and small in magnitude for the biology majors class (Table 5). There were no significant difference between pre- and posttest scores at the private research university and the midsized state college. Effect sizes for learning gains were negligible for these classes. It should be noted, however, that although students from the private research university did not demonstrate significant learning gains on the TOSLS over the course of the semester, they outscored all other classes on the pretest and posttest. It is also important to note that learning gains for midsized state college students may not be reflective of the gains possible across an entire semester, as the pre- and posttests were administered at midsized state college only 8 wk apart, as opposed to 16 and 14 wk between pre- and posttest administration at the public research university and the private research university, respectively. Consequently, our ability to compare learning gains from the midsized state college course with courses cross-institutionally is limited. These results may reflect differences in students’ scientific literacy development that are attributable to academic development, prior science learning, and student composition at different calibers of institutions.

### Table 6. Demographics of courses from each institution

<table>
<thead>
<tr>
<th>Public research university nonmajors</th>
<th>Public research university majors</th>
<th>Private research university nonmajors</th>
<th>Midsized state college nonmajors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project-based</td>
<td>Traditional</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male (% of sample)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major (% of sample)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social sciences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humanities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sciences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mathematics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Journalism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engineering</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undecided/not reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of college-level science</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>courses completed (average)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>290</td>
<td>296</td>
<td>544</td>
<td>50</td>
</tr>
<tr>
<td>38.3</td>
<td>26.4</td>
<td>40.1</td>
<td>32</td>
</tr>
<tr>
<td>3.49 (0.472)</td>
<td>3.53 (0.415)</td>
<td>3.27 (0.452)</td>
<td>3.62 (0.42)</td>
</tr>
<tr>
<td>3.5</td>
<td></td>
<td>36</td>
<td>15</td>
</tr>
<tr>
<td>1.2</td>
<td></td>
<td>12</td>
<td>3.8</td>
</tr>
<tr>
<td>78.1</td>
<td></td>
<td>40</td>
<td>21.3</td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td>0</td>
<td>12.5</td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td>2</td>
<td>17.5</td>
</tr>
<tr>
<td>4.6</td>
<td></td>
<td>0</td>
<td>3.8</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>6.25</td>
<td></td>
</tr>
<tr>
<td>1.46 (1.49)</td>
<td>1.12 (1.29)</td>
<td>2.43 (1.44)</td>
<td>2.48 (1.86)</td>
</tr>
<tr>
<td>0.53 (0.98)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We used an analysis of covariance (ANCOVA) to determine whether there was a significant difference among the three public research university classes in learning gains on the TOSLS, using pretest scores as the covariate. Data from the private research university and the midsize state college were excluded from this analysis, because the assumption of homogeneity of variance was violated when those data were included. Results from the ANCOVA yielded a significant main effect for class ($F = 11.380, p < 0.001$). Learning gains on the TOSLS were quantified by estimated marginal means (Weber, 2009). We chose to quantify learning gains using estimated marginal means rather than normalized learning gains, as the latter misrepresent or exaggerate gains when students are at extreme ends of the achievement spectrum in their pretest performance (Weber, 2009; Andrews et al., 2011). Estimated marginal means represent the estimated learning gains for each class after statistically controlling for the effect of pretest scores. These calculated means are appropriate given the ANCOVA. Post hoc pairwise comparisons with Bonferroni corrections indicated that the students in the project-based nonmajors class made significantly higher learning gains than the students in both the traditional nonmajors class and the biology majors class (Figure 4). There was not a significant difference in learning gains between the traditional nonmajors class and the biology majors class. It should be noted that students in the project-based nonmajors course demonstrated significant improvement from pre- to posttest on 10 of the 12 questions in which students from the traditional nonmajors course also made improvements (Figure 2). Additionally, students in the project-based nonmajors course made improvement on eight additional questions in skills 1, 4, and 6.

**DISCUSSION**

**Implications for Teaching and Learning**

Opportunities for developing skills such as argumentation and scientific reasoning are important and yet often missing from science education efforts (Newton et al., 1999; Norris et al., 2008; Osborne, 2010). We have developed this TOSLS instrument as a valid means to readily assess the impact of science, technology, engineering, and mathematics (STEM) education reform efforts on students’ development of these scientific literacy skills. In a recent survey of faculty, lack of support, articulated as not enough teaching assistants nor assessment tools, was identified as one obstacle to teaching science process skills (Coil et al., 2010). The TOSLS is a freely available, multiple-choice instrument that can be readily administered and scored in large-enrollment Gen Ed courses. The instrument contains items designed to specifically measure constructs related to using “evidence and data to evaluate the quality of science information and arguments put forth by scientists and in the media” (NRC, 1996). In particular, this instrument is responsive to the priorities of biology faculty, as the results from surveying biology faculty throughout the United States were critical to defining these skills. Instrument development was informed by relevant literature and multiple rounds of testing and revision to best reflect the common challenges in students’ development of scientific literacy skills.

An interesting finding that emerged in the process of developing the TOSLS is the disconnect between instructors’ value of scientific literacy, their teaching of these skills, and their assessment of students’ skill proficiency. More than 65.8% of faculty surveyed agreed that all nine skills were “important” to “very important” to scientific literacy. Similarly, most faculty reported that they teach and assess these skills (Figure 1; skills described in Table 2). However, when asked in an earlier open-ended question to state the three most important skills students need to develop for scientific literacy, many responses were related to biology content knowledge, rather than skills. This dissonance between what many faculty say they do and classroom reality has been documented by others and may be indicative of such concerns as the need to cover content and lack of time or expertise to develop and incorporate opportunities for skill development (Coil et al., 2010; Andrews et al., 2011; Ebert-May et al., 2011).

Because the TOSLS is sensitive enough to detect pre- to postsemester learning gains, its use may highlight the need to change or develop classroom activities that provide opportunities for students to develop the skills necessary to be scientifically literate citizens. This focus on developing scientific literacy skills is a major component in the push for reform in university STEM education, particularly in Gen Ed courses (Quitadamo et al., 2008; Chevalier et al., 2010; Coil et al., 2010; Hoskins, 2010). We used the TOSLS to evaluate the impact of a reformed Gen Ed biology course on student learning at a large public research university. Interestingly, we found that nonmajors students in our reformed classroom (project-based, Table 5 and Figure 4) made significantly greater learning gains than students in the traditional lecture-based course, even outperforming students in the biology majors course. Students in the project-based nonmajors course made greater gains than students in the traditional lecture-based course in several skill areas: skill 1 (question 1), skill 4 (questions 4 and 13), and skill 6 (questions 2, 6, 7, and 18) (Table 2). Students in the traditional nonmajors lecture-based course showed improvement in only two skill areas in which project-based students did not: skill 2 (question 22) and skill 9 (question 28). We are using the TOSLS to measure longer-term gains as we follow a subset of these students in subsequent courses.

We propose that instructors can use the TOSLS to identify the gap between their intentions to teach scientific literacy skills and students’ skill proficiency. In particular, using the TOSLS may spur greater alignment of learning objectives, classroom activities, and assessments. The TOSLS is
also informative in revealing student challenges and alternative conceptions in using scientific literacy skills. Instructors may use the TOSLS as a diagnostic tool in the beginning of the semester to reveal the extent of students’ literacy development. Class results may guide instructional planning. Additionally, instructors could tailor study suggestions for individual students’ skill development when using the TOSLS as a diagnostic assessment. An exploratory factor analysis indicates that the TOSLS instrument measures only one construct or trait rather than factors made up of our different scientific literacy skills, since one factor rather than two or more factors best accounted for the variance in the data. Common educational and psychological tests (Iowa Test of Basic Skills, Stanford Achievement Test) strive for unidimensional assessment (Osterlind, 2010). Because the instrument measures just this single construct, one can assume that responses of students to the test items reflect progress along a scale for scientific literacy. We envision that the TOSLS may be administered to inform classroom teaching and learning practices in a variety of ways. The TOSLS can be given, in its entirety, as an in-class pretest at the start of the course, using either the paper-based version or the Web-based version (currently in testing), and again as a posttest at the end of the course. Administration of the assessment in class is a means to motivate students to take the test seriously, and the variability in the amount of time spent completing the assessment is minimized.

Implications for Research
On the basis of our classroom testing of the TOSLS across course types and institutions, we expect that the TOSLS may serve as a useful assessment for other applications, including cross-institutional comparisons, evaluation of student learning over time, or as a means of programmatic assessment for Gen Ed curricula. However, comparisons between courses or different institutions are reliable only when the assessment is administered the same way. Further, the TOSLS items may be useful to instructors and other researchers to use as models to develop additional assessment items of their own. In particular, student challenges and misconceptions reviewed herein may be useful to inform additional assessment questions.

Limitations
Although we administered the TOSLS to a variety of students across multiple institutions, the test was developed using analysis of items administered to Gen Ed students attending a large research university in biology courses. The TOSLS shows promise for use across introductory undergraduate science courses, because our instrument-development process included alignment with STEM education policy guidelines; however, more research is needed to explore the validity of the instrument for use with science disciplines beyond biology (AAAS, 1993; National Academy of Sciences, 1997; Organisation for Economic Co-operation and Development, 2003; AAAS, 2010). Additional trials with the TOSLS may be warranted to fully clarify the utility of the test for different students under different situations. Many of the items required a degree of critical thinking and reading comprehension skills that may be lacking in some students; the lower proficiency observed in state college students may reflect this challenge. Alternatively, the lower gains observed in state college students may be indicative of the amount of time needed for students to develop skills between the pre- and posttest in order to observe gains. Finally, the observed gains in scientific literacy skills for Gen Ed and science majors at a large research university were not measured for time periods greater than one semester; longitudinal studies with these students could be very informative. Tests of the TOSLS under different situational factors may help address these questions.

Twenty years ago, educators could not have foreseen the rise of the Internet and the profound change in access to scientific information. Not surprisingly, most formal educational settings have lagged in integrating information evaluation criteria into existing curricula (Kuiper et al., 2005). The TOSLS questions were designed to address both these practical evaluation skills and scientific literacy skills needed by the general public. As new resources and access to scientific information change over time, policy documents inevitably follow with suggestions for incorporating these skills into educational settings. We hope that faculty will use this test to enhance how they respond to these recommendations.

The complete TOSLS is included in the Supplemental Material. We encourage instructors interested in using the TOSLS to contact the corresponding authors with requests for additional information. We also appreciate feedback on findings and comments for revisions for future versions of the TOSLS.

Institutional Review Board Protocols
Permissions to use pre- and posttest data and student demographics and to conduct student interviews, survey of biology faculty, and expert faculty evaluations were obtained (exempt, administrative review status: protocol nos. 2011-10034-0, -1, -2, -3) from the University of Georgia Institutional Review Board. Permissions to use pre- and posttest data and student demographics were obtained (protocol no. A0001392) from the Austin Peay Institutional Review Board. Permissions to use pre- and posttest data and student demographics were obtained (expedited status: protocol no. 201108240) from the Washington University in St. Louis Institutional Review Board.

ACKNOWLEDGMENTS
The authors acknowledge continuing support and feedback from the University of Georgia Science Educators Research Group. Diane Ebert-May, the external evaluator for a National Science Foundation grant-funded project supporting this work, provided critical comments throughout instrument development, as did Erin Dolan, Shawn Glynn, and Jenny Knight. Carly Jordan, Virginia Schutte, Sarah Jardeleza, and Greg Francom developed early versions of some instrument items and provided valuable commentary about items through the iterative process of revising items.

REFERENCES


The ability to interpret experimental data is essential to understanding and participating in the process of scientific discovery. Reading primary research articles can be a frustrating experience for undergraduate biology students because they have very little experience interpreting data. To enhance their data interpretation skills, students used a template called “Figure Facts” to assist them with primary literature–based reading assignments in an advanced cellular neuroscience course. The Figure Facts template encourages students to adopt a data-centric approach, rather than a text-based approach, to understand research articles. Specifically, Figure Facts requires students to focus on the experimental data presented in each figure and identify specific conclusions that may be drawn from those results. Students who used Figure Facts for one semester increased the amount of time they spent examining figures in a primary research article, and regular exposure to primary literature was associated with improved student performance on a data interpretation skills test. Students reported decreased frustration associated with interpreting data figures, and their opinions of the Figure Facts template were overwhelmingly positive. In this paper, we present Figure Facts for others to adopt and adapt, with reflection on its implementation and effectiveness in improving undergraduate science education.

INTRODUCTION

The ideal undergraduate biology classroom is a hub of active learning in which students engage in the investigative process of scientific research rather than memorizing facts (National Research Council, 2003; Handelsman et al., 2004; Labov et al., 2009; American Association for the Advancement of Science, 2011). Guided explorations of primary literature are especially useful in cultivating an inquiry-based learning environment and enhancing critical thinking skills (Muench, 2000; Levine, 2001; Smith, 2001). Primary research articles not only depict the nature of the scientific process, but also train students to evaluate data critically, improve their scientific writing, and follow the most recent advances in research (Levine, 2001; Kuldell, 2003; Hoskins et al., 2007). The acquisition of these and other information literacy skills increases students’ motivation to pursue research and prepares them to excel in graduate school, medical school, and research laboratories (Kozeracki et al., 2006; Hoskins et al., 2007).

Despite the importance of incorporating primary literature into the undergraduate biology curriculum, reading a scientific paper remains a daunting task for most students. The dense material, filled with unfamiliar terminology, technical details, and complex figures can easily overwhelm a novice reader. Difficult assignments that induce stress and anxiety in students can hamper the learning process (Pekrun et al., 2002). In our personal experiences, we have observed that introductory and upper-level students often approach a research article as they would a textbook, focusing on the narrative of the paper as fact, with the figures subordinate to the text. Undergraduates come to class having underlined and highlighted large portions of the text, yet they are often unable to describe the rationale for an experiment or
interpret data presented in the figures. Many students also fail to recognize that research articles often contain elements of persuasion and controversy and should be examined with critical eyes (Gillen, 2006). Thus, it is important that students engage the growing body of primary literature and learn specific strategies to analyze, interpret, and evaluate research papers critically.

A number of methods have been proposed and practiced to assist students with the analysis of primary literature (Janick-Buckner, 1997; Glazer, 2000; Levine, 2001; Gillen et al., 2004; Hoskins et al., 2007; Gehring and Eastman, 2008; Wu, 2009). A common classroom method is the journal club approach, in which one or more students present an article that has been read by the group (Glazer, 2000; Edwards et al., 2001; Kozeracki et al., 2006). A journal club approach, in which students become teachers, has the added benefit of sharpening students’ presentation skills, but the unstructured reading experience allows nonpresenters to engage the material passively and superficially. A second tactic often used by instructors is to design study questions to accompany each paper (Janick-Buckner, 1997; Levine, 2001; Wu, 2009). Study questions ensure that each student identifies the important points of the paper proactively, but this method is often time-consuming for the instructor and requires new questions for each paper, and questions must be crafted very carefully to prevent students from relying on the article text alone.

One particularly effective and novel approach to help students with primary literature is the CREATE method, described by Hoskins et al. (2007). CREATE utilizes a series of four related papers published by the same lab, which allows students to observe how the research process evolves over time. CREATE is unique in that it takes a data-centered approach to reading the papers; large sections of text are withheld from the students, and they answer figure-specific questions that require them to compare and contrast multiple individual panels. Students also predict future experiments and contact authors to gain a personal account of the research process. The in-depth CREATE approach requires a great deal of instructor preparation and relies on a nested set of sequential papers, which may not be practical for all course designs.

The Figure Facts template described here is similar to the CREATE method, in that it intentionally shifts students’ focus from the text to the data by supplying an analytical template with which to interpret each figure. However, Figure Facts is distinct from CREATE, in that Figure Facts takes a more streamlined approach to dissecting the article, often distilling the essential information into one or two pages. Importantly, the Figure Facts approach requires minimal instructor preparation, because there are no paper-specific study questions to write or grade, and the article is given to the students in its entirety, rather than being separated into discrete sections. The instructor-friendly Figure Facts method may be particularly useful for courses that utilize multiple, disparate primary literature papers on a regular basis. Figure Facts is flexible enough that students could easily adopt this method when reading primary literature in any course or scientific discipline. Figure Facts is not intended to replace other highly effective methods for teaching data analysis and interpretation, but may serve as an adaptable, complementary option for the instructor’s toolbox.

The primary goal of Figure Facts is to shift students’ focus away from the text of primary research articles and encourage them to spend more time interpreting the data figures. A secondary goal of the template is to provide a structured reading experience that reduces the frustration and anxiety that may serve as a barrier to student learning (Pekrun et al., 2002). To determine whether Figure Facts meets these educational objectives, the lead author (J.E.R.) implemented Figure Facts in an undergraduate cellular neuroscience course for two semesters. In the second semester, the authors measured whether students who used Figure Facts spent more time examining data figures in primary research papers. The authors also tested the assumption that a course based in primary literature improves students’ data interpretation skills. Finally, the authors examined students’ attitudes toward Figure Facts and determined whether this approach alleviates some of the frustration and anxiety that can reduce student learning (Pekrun et al., 2002).

**METHODS**

**The Figure Facts Template**

The Figure Facts template is a Microsoft Word table that students fill in electronically as they read a primary research article in preparation for class (Figure 1). The upper section of the Figure Facts template is devoted to the introduction of the paper. Students are asked to read the introduction carefully and fill in four key pieces of information that help them to identify the rationale for the experiments presented in the paper. First, they state the “broad topic” and the “specific topic” addressed by the research. These prompts serve to reinforce the relationship between the research paper and the current course subject matter, while simultaneously narrowing the focus of the paper. Next, the students identify “what is known.” Here, the students summarize previous findings that led to the current study. Listing prior knowledge is especially useful, in that it provides context for the paper and demonstrates continuity of the research process. Students should recognize that research builds on previous findings, and experiments are often based on unanswered questions generated by earlier studies. Finally, students state the “experimental question,” which prompts them to articulate the central question around which the paper’s experiments are based.

The main body of the template is devoted to the data figures of the primary research article. For each figure panel, students describe the specific experimental technique performed by the investigators. In the parallel column, students state the result and conclusion that can be drawn from each experiment. When filling out the template, students are instructed to dissect each figure before reading the corresponding text in the results section. They are asked to use their own words to describe the methods and results, rather than paraphrasing the figure legend or copying information from the text. Having students use their own words encourages students to interpret the data independent of the author’s interpretation. Students are also advised to use short phrases and abbreviations, rather than complete sentences, to minimize distraction from reading and understanding the paper itself.

**Implementation in the Classroom**

The lead author (J.E.R.) developed and used Figure Facts in two iterations of a Davidson College Cellular and Molecular Neuroscience course (BIO 333), which included...
14 biology and neuroscience majors in Spring 2011 and 16 students of a similar demographic in Spring 2012. The elective course consists of two 75-min gatherings per week for 16 wk. Each class period is a fluid mixture of lecture, discussion, and data analysis. All reading assignments come from primary literature, scholarly reviews, and online resources. The primary research articles usually contain microscopic images and/or complex graphical representations of averaged data sets. When reading these articles in preparation for class, students are required to complete a Figure Facts template (Figure 1). Students type their responses to the prompts in the electronic document using word-processing software.

Prior to the class meeting, students upload the completed template into an assignments folder in Moodle, a course-management system. Uploading Figure Facts prior to class requires students to make a significant effort to understand each figure before the group meeting, rather than skimming the paper briefly before class. The upload requirement also follows the principles of just-in-time teaching (Novak et al., 1999), in that it allows the instructor to assess each student’s level of understanding prior to class and adjust group discussion to suit the students’ needs.

During class, each figure is projected using Microsoft PowerPoint, and we examine the data closely as a group. With PowerPoint, the instructor can divide more complex figures into multiple parts, which prevents students from becoming overwhelmed by too much information at once. Projecting the figures guarantees that everyone is looking at the correct figure panel and allows students to interact with the data using laser pointers that are passed around the room.

The instructor asks questions to structure the conversation, but the students are expected to articulate the methods and results of each figure, identify weaknesses in experimental design, and propose future experiments. Students are instructed to bring a hard copy of their completed template to class, so they can refer to their notes as we examine the data. Students are encouraged to write on the template and fill in any missing information as we go along. Later, the template serves as a condensed study guide to aid the student in preparing for quizzes and exams. Collectively, Figure Facts are worth 10% of each student’s course grade, which provides sufficient incentive for students to complete each assignment. To minimize the time required to grade the templates, we scored them on a scale of 1–5. While the students are not expected to interpret every figure correctly, they are asked to provide evidence that they made a significant effort to understand each experiment. No part of the template can be left blank, and each box has to contain at least a best guess or a question related to the figure panel. With a cursory examination of each template, the instructor is able to determine whether the information is complete, contains sufficient detail, and is written in the student’s own words. On average, each template should take only 1 or 2 min to examine and grade. The first template of the semester is scored but not included in the grade calculation, which allows students to become familiar with expectations and receive feedback from the instructor.
The primary goal of Figure Facts is to shift students’ focus away from the text of research articles and encourage them to spend more time examining the figures, thus sharpening students’ data interpretation skills. A secondary goal of Figure Facts is to provide a structured reading experience that will reduce some of the frustration and anxiety students often experience when learning to read primary literature. To determine whether Figure Facts accomplishes these educational objectives, we designed a three-part assessment plan. First, to determine whether Figure Facts increases students’ attention to data figures, we performed a time-on-task assessment, in which students documented their paper-reading activities before and after intervention with Figure Facts. Specifically, in weeks 1 and 15 of the Spring 2012 semester, students recorded the amount of time they spent reading each section of a primary research article. They were asked to differentiate between the time they spent examining data figures and the time they spent reading about those figures in the text. We also recorded the number of handwritten notations that students made in the margins of each figure panel to ascertain the levels of student engagement with the visual data. It is important to note that students were not required to fill out a Figure Facts template for the week 1 or the week 15 assessment. This element of the assessment design allowed us to determine whether students spent a greater percentage of time examining figures after using Figure Facts for one semester, even in the absence of the template. Student data from weeks 1 and 15 were analyzed statistically using a paired t test.

Second, to determine whether a course grounded in primary literature sharpens students’ data interpretation skills, we administered three ungraded skills tests at regular intervals during the Spring 2012 semester. (The skills tests are available upon request and are appropriate for biology majors with prior basic knowledge of neuron structure and function and previous experience interpreting graphs.) Each 30-min test consisted of two figures containing microscopic images of neurons and graphs of averaged data sets. Students were asked to examine the figures and then identify the true statements from a list of possible conclusions. All students had completed the prerequisite course, BIO 111 (Molecules, Genes, and Cells), in which they discussed neuron structure and function, examined microscopic images, and interpreted graphs. Therefore, each student had sufficient content knowledge to complete the assessment. Each test contained 14 answer choices and 5 correct answers, yielding a score range of 0 to 14. The tests were administered during class time to maximize students’ attention and effort. Student performance was tracked using anonymous identification numbers, and data were analyzed using paired t tests.

We also administered the skills tests to a control group of 15 Davidson students who had completed the prerequisite course, BIO 111, and thus had experience with neurons, microscopic images, and graphs. Students who completed the tests in identical order in one sitting showed no significant improvement between the first and second test (t(13) = 0.18, p = 0.86) or the first and third test (t(13) = 1.94, p = 0.08). These control data suggest that the three skills tests are comparable with regard to difficulty, and any improvement exhibited by BIO 333 students over the course of the semester is not attributed to a more difficult first test.

Third, to assess students’ opinions of the Figure Facts approach, we administered anonymous midsemester surveys in both iterations of the course. The midsemester survey utilized a 5-point Likert scale, which asked students to agree or disagree with statements pertaining to the use of Figure Facts. In Spring 2012, we added pre- and postcourse surveys to assess students’ experiences and attitudes toward reading primary literature in general. Specifically, we used a 5-point Likert scale to ask students whether they considered reading primary research articles to be a frustrating and/or stressful experience, and we examined whether a semester of guided literature exploration reduced negative emotions associated with this complex task. The overall IRB study participation rate for the three-part assessment was 100%, with some student scores and opinions omitted due to absence from class or incorrect completion of the assignment. Students consented to be included in the study by completing and submitting the precourse survey described above.

**RESULTS AND CONCLUSIONS**

**Time-on-Task Assessment**

The primary goal of Figure Facts is to shift students’ focus away from the text of primary research articles and to encourage students to spend more time interpreting the data figures. To determine whether Figure Facts was associated with increased student attention to data interpretation, we performed a time-on-task assessment in which students documented their paper-reading activities before and after exposure to Figure Facts. In week 1, before students were introduced to the Figure Facts concept, students spent approximately 80% of their time reading the text and 20% of their time examining the data figures (Figure 2A). In week 15, after completing eight Figure Facts templates, students spent only 60% of their time reading the text, and doubled their time examining data to 40%. This substantial increase in the percent time spent interpreting data figures was statistically significant (t(13) = 5.5, p < 0.001).

In addition to tracking data interpretation as a percentage of time, we recorded the number of handwritten notations that students made in the margins of each figure panel (Figure 2B). No instructions or grades were given for handwritten notes on the paper, so the students had no external incentive for making these notations. In week 1, students made notations on 10% of the printed figure panels, whereas in week 15, students marked 38% of the figure panels (t(15) = 4.4, p < 0.001). Therefore, students appeared more actively engaged in examining figures for comprehension in week 15, as indicated by the nearly fourfold increase in handwritten notations on hard copies of the research papers. It is important to note that students were not required to complete a Figure Facts template for the week 15 assessment. Therefore, prior experience with the template was sufficient to induce a significant and measurable shift in students’ paper-reading habits.

**Data Interpretation Assessment**

An underlying assumption of the Figure Facts approach is that students who practice reading primary research articles for comprehension will sharpen their data interpretation...
students who examine data figures repeatedly will improve their ability to comprehend complex graphical representations of data and draw accurate conclusions. To test this assumption we administered three skills tests at regular intervals throughout the semester. Each test consisted of two figures containing microscopic images of neurons and graphs of averaged data sets. Students were asked to examine the figures and identify true statements from a list of possible conclusions. In week 1, the average student score was −1 (on a scale of −14 to +5), indicating that most students tended to choose false conclusions more often than they chose correct conclusions (Figure 3; n = 16). In week 9, after students had analyzed eight primary research papers in detail, the average score improved significantly (t(14) = 4.22, p < 0.001) to +1. By week 15, students’ performance trended upward to +2, but the difference between week 9 and week 15 was not statistically significant (t(1) = 0.94, p = 0.36). The measurable improvement observed from week 1 to week 9 suggests that regular, structured primary data analysis improves students’ abilities to interpret novel data sets.

Figure 2. Students spent more time examining data figures after using Figure Facts for 15 weeks. (A) In weeks 1 and 15, students recorded the amount of time they spent reading the text of a primary research article vs. examining the data figures. (B) The number of notations made in the margins of each figure panel was also recorded at weeks 1 and 15. Error bars, SEM; ***, p < 0.001 by paired t test; week 1: n = 14; week 15: n = 16. Two students were omitted from the week 1 analysis due to incorrect completion of the assessment.

Figure 3. Students interpreted novel data sets more accurately at week 9. At weeks 1, 9, and 15, students were asked to examine a collection of microscopic images and graphs and identify true statements from a list of possible conclusions. Positive values indicate more correct than incorrect answers. Box represents the middle 50th percentile, line represents the median, and whiskers represent minimum and maximum values. ***, p < 0.001 by paired t test; n = 16.

Student Attitude Surveys
To assess students’ attitudes toward Figure Facts and primary literature in general, we asked students in both iterations of the course to complete anonymous midsemester surveys. Eighty percent of the students agreed or strongly agreed that Figure Facts helped them structure their reading of primary research papers (Figure 4A). One student remarked that dividing the readings into discrete, manageable steps reduced some of the anxiety associated with tackling unfamiliar research papers. Ninety percent of students agreed or strongly agreed that Figure Facts helped them focus on the data, rather than the text of the paper. One student wrote the following comment on the anonymous midsemester survey:

I really appreciate the Figure Facts ... I used to skip over the figures in articles because they were too difficult to understand, but forcing me to focus on the data is teaching me a better way to read and understand research.

This comment indicates that Figure Facts generates the desired educational goal, which is to guide students away from their dependence on the text and encourage them to persist in their efforts to interpret the data.

The midsemester surveys also addressed issues of class preparation and discussion. Eighty percent of students agreed or strongly agreed that Figure Facts assignments encouraged them to read the paper more closely in preparation for class. Requiring students to upload their template and assigning a grade to their effort encouraged students to regard the paper analysis as a high-priority exercise, rather than a passive reading assignment that they could skim right before class. Sixty percent also agreed or strongly agreed that the template helped them organize their thoughts during class discussion. Long moments of silence during discussion times were rare, because students could easily refer to their prepared comments if they had trouble recalling information pertaining to the figure at hand. Students were able to recall basic information quickly, which gave us more time to discuss the benefits and drawbacks of each experiment, critique the experimental design, and propose improvements.
to the study. Finally, 80% of students agreed or strongly agreed that the template made our classroom discussions more productive, suggesting that a well-prepared and well-organized approach to discussing primary literature in the classroom is favored by instructors and students alike.

In Spring 2012, we administered pre- and postcourse surveys to assess the levels of stress and anxiety students experience when reading primary literature. In week 1, 47% of students agreed or strongly agreed that reading primary research papers was often a stressful experience for them, and 58% reported that they became frustrated when trying to understand papers (Figure 4B). By the end of course, only 12% of students agreed that reading research papers was a stressful experience and only 19% reported feeling frustrated (Figure 4B). These findings indicate that repeated exposure to primary data, combined with a structured reading approach, is associated with reduced student anxiety and frustration. Given that heightened stress and anxiety correlates inversely to learning gains (Pekrun et al., 2002), reducing frustration may aid students significantly in learning the data interpretation skills that are central to evaluating research papers. This finding complements a recent report by Hoskins et al. (2011), which shows that guided literature explorations increase students’ overall confidence in their ability to interpret data.

At the end of the semester, students also completed standard, college-issued course evaluations, which asked students to comment on specific elements of the course that they found helpful to the accomplishment of the course’s goals, as well as areas for improvement. Half of the students spontaneously provided positive comments pertaining to Figure Facts in this section of the form (Table 1), and no negative comments pertaining to Figure Facts were encountered. These unsolicited comments suggest that many students maintained a favorable opinion toward Figure Facts at the end of the semester and regarded the innovation as a highlight of the course.

**DISCUSSION**

**Future Improvements**

The preliminary use of Figure Facts in the classroom was successful in that students shifted their focus from the text of the articles to the data figures (Figure 2), improved their ability to interpret complex data sets (Figure 3), and experienced

---

**Table 1.** Selected student responses to the following end-of-semester course evaluation question: What elements of the course and instruction did you find most helpful to the accomplishment of the course’s goals?

- “Figure Facts was a great teaching tool for article comprehension.”
- “Figure Facts were helpful in learning how to read modern science writing.”
- “Figure Facts were helpful in forcing us to finally analyze papers in detail.”
- “Figure Facts kept me up-to-date and organized.”
- “Figure Facts were helpful in understanding the primary sources.”
- “Figure Facts actually really helped me improve my scientific reading abilities.”
- “Figure Facts were challenging, but a really helpful and unique way to learn about experiments, and a great way to study.”

---
reduced frustration and anxiety with regard to reading primary literature (Figure 4). However, it is interesting to note that, on the midsemester survey, only 38% of students agreed that Figure Facts helped them to identify how the study related to previous research (Figure 4A). From this finding, we conclude that the “what is known” prompt in the table is not effective in helping students to understand how the paper relates to previous findings. Perhaps simply changing the name of this prompt from “what is known” to “previous research” may make the link between the current paper and earlier experiments more obvious. Student feedback also indicated that we should spend more time discussing the previous findings at the beginning of class in order to place the current experiments into context. It may be especially useful to assign two or more related papers, as advocated by the CREATE approach (Hoskins et al. 2007), to better illustrate the continuity of the research process.

A second potential improvement to the Figure Facts template would be to move the “experimental question” prompt to the bottom of the template. After spending a great deal of time on the details of each figure, it may be useful for students to step back, state the main conclusion of each figure, and confirm the experimental question being addressed by the authors. By stating the main idea at the end of the discussion, students may be able to articulate the take-home message of the paper more clearly and gain a greater appreciation for the complex series of experiments required to generate evidence in support of a single hypothesis. It may also be beneficial to ask, “What is the next experiment?” at the bottom of the Figure Facts template to encourage students to think about appropriate future experiments in preparation for group discussion.

An important caveat of this study is that it was performed with small, highly selected samples of students who had close interaction with a single instructor. Student and instructor motivation may influence positive assessment gains and student attitudes to some degree, so the effectiveness of Figure Facts should be assessed on a larger scale in different classroom and campus settings before generalizations can be made. The results of the data interpretation skills tests should also be regarded as preliminary, due to the small scale of the assessment. Science educators could benefit greatly from a larger-scale study examining the impact of primary literature exposure on students’ ability to interpret complex data sets. Finally, while the overwhelmingly positive student feedback highlighted in Figure 4 and Table 1 is encouraging, it would be interesting to determine whether students are inclined to use Figure Facts on a voluntary basis, both in the classroom and in an independent research setting.

**Summary**

The Figure Facts template provides a structured reading exploration for undergraduate students as they learn how to analyze primary literature. We found that Figure Facts does indeed increase students’ efforts to interpret data figures, while simultaneously reducing the frustration often associated with this task. With this increased attention to data figures, students improved their data interpretation skills when presented with novel data sets. With basic figure content mastered in advance, students could devote more class discussion time to critical analysis of experimental design, alternative interpretations of results, and potential future experiments. Figure Facts can be adapted easily to any research article in any science course, and is one of many tools that instructors may use to demonstrate the investigative nature of biology through primary literature.

**ACKNOWLEDGMENTS**

We thank students in the Spring 2011 and 2012 Cellular and Molecular Neuroscience courses at Davidson College for their enthusiasm, participation, and feedback regarding the use of Figure Facts in the classroom. We also thank students in the Fall 2012 Stem Cell Biology course for serving as a control group in our assessment. We thank Barbara Lom for thoughtful comments on the manuscript and assessment design, and we are grateful to Patricia Peroni for consultation regarding statistical analysis. J. Round is supported by an Undergraduate Science Education Grant from the Howard Hughes Medical Institute, a Research at Undergraduate Institutions Award (1049768) from the National Science Foundation, and Davidson College.

**REFERENCES**

American Association for the Advancement of Science (2011). Vision and Change in Undergraduate Biology Education: A Call to Action, Washington, DC.


Wu J (2009). Linking assessment questions to a research article to stimulate self-directed learning and develop high-order cognitive skills in an undergraduate module of molecular genetics. CBE Life Sci Educ 8, 283–290.
CREATE Cornerstone: Introduction to Scientific Thinking, a New Course for STEM-Interested Freshmen, Demystifies Scientific Thinking through Analysis of Scientific Literature

Alan J. Gottesman and Sally G. Hoskins

City College and the Graduate Center, City University of New York, New York, NY 10031

Submitted November 21, 2012; Revised December 15, 2012; Accepted December 15, 2012

Monitoring Editor: Jose Vazquez

The Consider, Read, Elucidate hypotheses, Analyze and interpret data, Think of the next Experiment (CREATE) strategy for teaching and learning uses intensive analysis of primary literature to improve students’ critical-thinking and content integration abilities, as well as their self-rated science attitudes, understanding, and confidence. CREATE also supports maturation of undergraduates’ epistemological beliefs about science. This approach, originally tested with upper-level students, has been adapted in Introduction to Scientific Thinking, a new course for freshmen. Results from this course’s initial semesters indicate that freshmen in a one-semester introductory course that uses a narrowly focused set of readings to promote development of analytical skills made significant gains in critical-thinking and experimental design abilities. Students also reported significant gains in their ability to think scientifically and understand primary literature. Their perceptions and understanding of science improved, and multiple aspects of their epistemological beliefs about science gained sophistication. The course has no laboratory component, is relatively inexpensive to run, and could be adapted to any area of scientific study.

INTRODUCTION

We think a significant number of students lose interest in studying science early in their college careers, because many science curricula do not promote open-ended discussion, critical analysis, and creative study design—activities that characterize science as it is practiced. We thought that one way to attract and retain students who might be considering science studies would be to give them an opportunity to develop their reading and analytical skills and gain a realistic sense of scientific thinking as soon as they started college. A Consider, Read, Elucidate hypotheses, Analyze and interpret data, Think of the next Experiment (CREATE)-based course focused on scientific thinking, using a novel selection of readings whose analysis did not require years of content mastery, would, in principle, give freshmen a chance to engage deeply in activities characteristic of actual science practice. We hypothesized that such an experience could have a positive influence on students’ scientific abilities, their attitudes toward science, and their understanding of the research process early in their academic careers. To test this idea, we developed a new elective, Biology 10050: Introduction to Scientific Thinking.

BACKGROUND

Biology 10050 was developed as an adaptation of an upper-level course, Biology 35500: Analysis of Scientific Literature proposed by Hoskins et al., 2007.
with CREATE. That course, offered at City College of New York (CCNY) since 2004, aims to demystify and humanize science through intensive analysis of primary literature. In Biology 35500, “modules”—sets of journal articles published sequentially from single laboratories—are the focus for an intensive elective. Students learn a new set of pedagogical approaches, including concept mapping, cartooning of methodology, figure annotation, use of templates to parse experimental logic, and design of follow-up studies (Hoskins and Stevens, 2009; Hoskins, 2010b). These methods are applied first to an article from the popular press and then in the analysis of a series of primary literature papers that follow a particular scientific question (e.g., “How do axons find their targets in the embryo?” “How is axis polarity maintained during regeneration?”). By examining module articles in a stepwise manner, we develop a “lab meeting” atmosphere in the class, with experimental findings discussed as if they had been generated gradually by the students themselves. Within individual articles, every figure or table was analyzed with recognition that each specific question being addressed or question asked created a data subset that contributed to the major finding of the paper.

In CREATE class sessions, multiple aspects of study design are scrutinized closely as we work backward from data in each figure and table to reconstruct details of the particular experiment that generated those data before we analyze the findings. In the process of examining specific experiments and their outcomes, we repeatedly consider questions fundamental to much research, (e.g., “What is n?” “How was the sample selected?” “What controls were done and what did each control for?” “How do the methods work?” “What is the basis of ‘specificity’ in staining, binding, or expression?” “How convincing are the data?”). In addressing such questions, students gain insight into the design and interpretation of research beyond the individual study under analysis. Because methods are examined in terms of fundamental biological and chemical properties (e.g., “What makes antibodies ‘specific’?” “Do antibody probes bind the same way that riboprobes do?” “How can you tell whether a particular stem cell undergoes division after injury to an organism?”), students review fundamental content from previous course work in a new context. By considering “evolution of methodology” (e.g., differential screening of CDNA libraries vs. gene chip analysis vs. RNAseq approaches; gene knockout vs. RNA interference) students become aware of the pace of technique development and how the range of tools available may influence the nature of questions asked. In this way, Biology 35500, the original CREATE course, involves both close analysis of papers presented in their original sequence as an individual “module” but also consideration of broader nature of science issues. For example, discussion centered on the fact that what used to be considered “junk” DNA is now recognized as having a key role in microRNA pathways illustrates the malleability of scientific knowledge.

After completing analysis of each paper, and before moving to the next paper in the series, students create their own follow-up experiments, thereby building experimental design skills, as well as awareness that a given study could, in principle, move forward in a variety of ways. Students’ proposed follow-ups are vetted in a grant panel exercise designed to mimic activities of bona fide panels (see Hoskins et al., 2007). In turn, these sessions lead to discussion focused on broader scientific issues, including interlaboratory competition, peer review, and the factors that might influence principal investigator (PI) decisions about what direction to take next.

Late in the semester, students, as a class, develop a list of 10–12 questions for paper authors. These are emailed as a single survey to each author (PIs, postdocs, graduate students). Many authors reply with thoughtful comments about their own paths to science, their motivations, and their lives beyond the laboratory. Discussion of authors’ varied responses complement the in-class data analysis with insight into the lives and motivations of “the people behind the papers.”

Our upper-level course led to gains in students’ content integration and critical-thinking ability, as well as in their self-assessed learning gains (Hoskins et al., 2007). We also found that undergraduates’ self-assessed science abilities, attitudes, and epistemological beliefs changed during the CREATE semester (Hoskins et al., 2011). Upper-level students’ postcourse interviews (see Tables 1 and S1 in Hoskins et al., 2007), as well as conversations with alumni of Biology 35500 (“You have to do a version of this for freshmen—it changed how I read everything” and “If I had known sooner that research wasn’t boring, I might have joined an undergrad research program”) inspired us to consider adapting upper-level CREATE for freshmen.

A related motivation for developing the CREATE Cornerstone course was that the biology department at CCNY, like its counterparts elsewhere, loses many would-be majors during the early years of the biology curriculum. Some students who start with the intention of declaring a biology major do not follow through. Others who do choose biology later change majors and leave science altogether, with multiple factors likely playing a role. Students may be poorly prepared for college-level science, feel overwhelmed by the amount of new information covered in the introductory-level courses (Seymour and Hewitt, 1997), or be discouraged by textbooks’ depiction of biology as a largely descriptive science (Duncan et al., 2011). Nationwide, some students get the impression from the laboratory components of introductory biology, chemistry, or physics classes that lab work is routine, predictable, and boring.

We felt that a CREATE Cornerstone course focused on scientific thinking could support and build students’ science interest at an early phase of their academic careers. In part, adapting upper-level CREATE for freshmen might benefit students by teaching them a variety of techniques (the CREATE toolkit; Hoskins and Stevens, 2009) that make complex material more accessible and understandable. At the same time, the course seeks to provide students with an inside look at the workings of real-world biology research labs and the diversity and creativity of the scientists who work in them. We hypothesized that students in such a course would become more adept at thinking critically about scientific material and at designing and interpreting experiments—key strategic foci of the CREATE approach. In addition, we hypothesized that students would gain in their abilities to critically analyze scientific writing, deepen their understanding of the nature of science, and develop more mature epistemological beliefs about scientific knowledge. We also suspected that some students who had not considered careers in research, or others who had but quickly rejected the idea, would consider research more positively as their college education progressed.
Introduction to Scientific Thinking is a three-credit, one-semester elective for first-year college students with a declared interest in science, technology, engineering, and math (STEM) disciplines at the CCNY, a minority-serving institution. The course meets twice-weekly for 75 min/session, and on our campus is taken before the introductory-level courses in any of the basic sciences. The goal is to develop the science-related reading and analytical skills of freshmen by using the CREATE strategy to critically evaluate a number of recent and ongoing research studies. Ideally, the experience should also encourage students to persist in STEM disciplines, participate in undergraduate research experiences (UREs) in later years, and consider research as a career choice.

At CCNY, first-year students cannot declare a biology major. The course is thus aimed at presumptive biology majors and in principle could be taken concomitantly with the standard introductory biology (or other science) course. On campuses where students can or must declare a major in the first year, this course would be appropriate for students who evince interest in biology studies. The data reported here address changes in Biology 10050 students' critical-thinking/experimental design abilities and in their attitudes and beliefs about science. The question of student persistence in STEM and participation in undergraduate research projects will be tracked in upcoming semesters.

METHODS AND ASSESSMENT TOOLS

Participants in this study were first-year students at CCNY who enrolled in the semester-long Biology 10050: Introduction to Scientific Thinking course during Fall 2011 and Spring 2012. In each semester, at the first class session, students were invited to participate anonymously in our study on a voluntary basis that had no bearing on class grade. Precourse data were collected during the first few classes and postcourse data in the final class session of the semester. All participating students were asked to devise a “secret code” number known only to them and to use this code on all surveys. Identifying students were asked to devise a “secret code” number known only to them and to use this code on all surveys. Identifying surveys in this way allowed us to compare individual and group scores pre- and postcourse, while preserving student anonymity (Hoskins et al., 2007).

Critical Thinking Assessment Test (CAT). Students in the Fall cohort of Biology 10050 completed the CAT (Stein et al., 2012). In the CAT, which is a reliable and valid test of critical thinking, students spent 1 h reading a number of informational passages and writing responses to a variety of prompts asking them to evaluate the information and draw conclusions. The same test was taken again at the end of the semester. The CAT tests were graded and analyzed statistically (Student’s t test) by a scoring team at Tennessee Tech University, where this survey was created.

Experimental Design Ability Test (EDAT). Students in both cohorts of Biology 10050 also completed the EDAT, the reliability and validity of which have been established by the EDAT developers (Sirum and Humburg, 2011). In the EDAT, students were presented with a claim and challenged to “provide details of an investigative design” and indicate the evidence that would help them decide whether to accept the claim. Students were given 15 min to respond to a written prompt that described the assertion. Precourse and postcourse versions of the EDAT present different scenarios. Precourse, students read a paragraph presenting the claim that the herb ginseng enhances endurance; postcourse, the selected text alleged that iron supplements boost memory. The EDAT survey was scored separately by two investigators following the scoring rubric created and explained in Sirum and Humburg (2011). After the individual scoring, any discrepancies were discussed and reconciled. Tests for statistical significance were performed using the Wilcoxon signed-rank test (http://vassarstats.net/index.html; Arora and Malhan, 2010). Effect sizes (Cohen, 1992; Coe, 2002) were also determined.

Survey of Student Self-Rated Abilities, Attitudes, and Beliefs (SAAB). To investigate students’ reactions to the CREATE course, we asked them to complete the SAAB. In this Likert-style survey, students reported their degree of agreement on a 5-point scale (range: strongly disagree to strongly agree) with a series of statements concerning their attitudes, self-rated abilities, and beliefs about analyzing scientific literature; the research process; the nature of scientific knowledge; and scientists and their motivations. The surveys were identical precourse and post course, and used statements whose derivation and description is described in Hoskins et al. (2011). Students were given 20 min to complete the survey. For statistical analysis, all response scores were aggregated into their appropriate categories (see Supplemental Material for derivation of categories) and changes precourse to postcourse were analyzed for statistical significance using the Wilcoxon signed-rank test. Because these data and those of the EDAT are nonparametric (a score of “4” is not twice as good as a score of “2,” for example) and noncontinuous, the signed-rank test was deemed an appropriate analytical tool (Arora and Malhan, 2010).

The SAAB data for the Biology 10050 class include pooled results from the Fall and Spring sections (18 and 13 participating students, respectively). Data collected using the same survey, administered in the same manner, were also obtained from one contemporaneous section of the upper-level CREATE course (Biology 35500, 21 students; two meetings per week for 100 min/session). Additionally, the SAAB survey was administered to volunteers in a course in Organismic Biology (general physiology, 23 students; one 100-min lecture and one 3.5-h lab session/wk), none of whom had taken a CREATE class. This group was not a matched-control population (students were not freshmen). Rather, data from this cohort of students provided insight into potential changes in attitudes, abilities, and epistemological beliefs that might happen naturally during the course of a semester in a nonCREATE science class. The CREATE classes were taught by the same instructor (S.G.H.); the Organismic Biology class was taught by a colleague not otherwise involved in this study. Both instructors were experienced at teaching their respective courses.

Student Comments on Author Emails. To gain insight into students’ reactions to author email responses, we assigned students to read and annotate the responses as they were received. Students included the responses in their notebooks/portfolios, with marginal notes indicating which aspects of each response they found most surprising and/or interesting. In the Spring session of Biology 10050, we included a question on a late-semester (in-class, open-book) exam,
asking students whether the emails changed their ideas about science research or scientists. We compiled responses and analyzed them for repeated themes.

**Student Participation.** The CREATE study was approved by CUNY Institutional Review Board (Exemption category 1 and 2). Of the students in Bio 10050, 69% were female and 59% were members of minority groups currently underrepresented in academic science. Students were invited, in week 1 of class, to anonymously participate in an education study with the goal of “improving undergraduate education in science.” Participation was optional and the instructor noted that student participation or nonparticipation had no bearing on course grade or any other relationships with CCNY. There were no points or extra credit awarded for participation. We think that students who participated were motivated by the chance to take part in a science education study and/or to be part of a scientific experiment.

**CURRICULAR DESIGN**

**Adapting CREATE for Freshmen**

In the original (upper-level) CREATE course, the class studied, sequentially, a series of papers published by a single lab that tracked the development of understanding in a particular field of scientific inquiry (e.g., how embryonic retinal axons find their targets in the brain; how planaria maintain positional information during regeneration). For the freshmen, we changed the types of articles studied, using popular press articles and a wider range of scientific literature, but applied the same overall CREATE teaching/learning strategies. The freshmen initially read and analyzed numerous popular press stories based on journal articles. We also read a variety of newspaper and magazine pieces describing scientific investigations or researchers. These warm-up exercises, used more extensively for the freshmen than in upper-level CREATE, started students toward developing the skills they would need for reading and analyzing primary literature later in the semester. All the readings (in all CREATE courses) are actual texts as originally published. In some cases, we read only parts of papers, but we did not rewrite or simplify any of the material. The freshmen ultimately read a pair of papers published in sequence that addressed a subject—the ability of infants to recognize and judge the social actions of others—related to a number of the shorter readings.

Toward the end of the semester, the freshmen, as a class, composed a list of 10–12 questions about the studies we had read, “research life,” and the researchers themselves. These questions were emailed as a single survey to each paper’s authors, with a cover letter explaining our approach and inviting a response. This key strategic component of CREATE courses seeks to shift students’ often-negative preconceptions about what research/researchers/research careers are like. Many of the scientist-authors responded with comprehensive answers related to their personal and professional lives, their contributions to the work that we studied, and their scientific experiences as their careers developed. The generosity of authors in preparing thoughtful responses is especially valuable and memorable, according to our students.

**CREATE Cornerstone Objectives and Selected Exercises**

Students learned to use CREATE tools, including concept mapping, paraphrasing, cartooning, annotating figures, applying templates to parse experimental logic, designing follow-up experiments, and participating in grant panels (Hoskins and Stevens, 2009). The CREATE techniques aim to sharpen students’ analytical skills and build critical-reading habits that can be used in new situations. These approaches also build students’ metacognition—the ability to track their own understanding (Tanner, 2012). To construct a concept map successfully, for example, students need to understand individual ideas and discern the relationships between them. To sketch a cartoon that shows what took place in the lab to generate the data presented in a particular figure, students must make sure they understand the relevant methodology. We applied concept mapping and cartooning along with other CREATE tools to a novel combination of readings. Articles selected for Biology 10050 were chosen because of their topicality, relatively simple methodology, and aspects of each that provoked controversy, exemplified the role of controls, and/or highlighted important distinctions between data and their interpretation. Goals for the Cornerstone students included learning: to read with skepticism, to critically analyze data and generate alternative interpretations, to recognize the malleability of scientific knowledge, and to develop and evaluate experiments with particular emphasis on controls and their roles. A final goal was for students to develop a more realistic view of research and researchers than the one often promoted in popular culture.

**Developing an Appropriately Skeptical Reading Style**

The class sessions were typically run as discussions or debates about points that arose in the assigned readings. We rarely presented all the information at once, instead examining each reading in stages. For example, one unit early in the semester used an op-ed in the *New York Times* claiming that iPhone owners experienced “love” for their phones and outlining study outcomes that purported to support this conclusion (Lindstrom, 2011). We also read a published refutation of the op-ed signed by 44 neuroscientists (Poldrack, 2011a), and the original version of the refutation letter before it was edited by the *New York Times* (Poldrack, 2011b). We started with the op-ed and only later distributed the challenge from the neuroscience community, considering:

- How, in principle, would one determine “the most appealing sounds in the world,” whether babies “automatically” swipe iPhones expecting a response, or whether “love” is experienced by phone owners (as claimed by Lindstrom, 2011)?
- What evidence would you find convincing?
- What studies would you do if you were interested in such issues?
- How did Lindstrom make such determinations?
- On what basis do the neuroscientists challenge the stated conclusions?
- Do the *New York Times*’ edits shift the message of the original letter to the editor? If so, how?
- Taking all of the readings and analyses together, what do you conclude about iPhone “love”? Why?
As they learned to use and apply CREATE tools, students accustomed to reading and passively accepting the information encountered in their textbooks, on the Internet, or in newspapers began to recognize that just because something is published does not mean it is beyond criticism (Hoskins, 2010a).

Data Analysis—Developing Alternative Interpretations

“Writing about Testing Worries Boosts Exam Performance in the Classroom” (Ramirez and Beilock, 2011) is a Science paper examining the degree to which stress may contribute to undergraduates’ “choking” on exams. We initially distributed only some of the paper’s narrative and a single figure illustrating the first study performed, holding back the title, abstract, and all other information. During class, students diagrammed the experiment, which compared test scores of two groups of students. Each group had been administered a baseline math test. Posttest, both groups were told a stress-inducing story about how outcomes on a later test covering the same material would be used. Before taking the second test, one group wrote for 10 min about their fears of poor test performance, while the other group sat for 10 min. The data revealing the test scores of the two groups show the nonwriting group performing worse on the second test than they did on the first, thus “choking,” while the writing group scored gains. We considered:

Can we conclude that writing about one’s test concerns leads to less choking on exams? How solid is that conclusion? If we had generated these data ourselves, could we publish now? Why? Why not?

Are any alternative interpretations of the data plausible?

Through discussion, students proposed a third “write about anything” group as an additional control. We next provided the paper’s figure 2 and associated narrative. The authors had added a third group that was instructed to write about “an unrelated unemotional event.” Students saw that the investigators had added the same control group they had asked for, extending the study to resolve the “writing-only” issue. This bolstered students’ sense that they were “thinking like scientists.”

Using Sketching to Clarify Design—Developing Alternative Interpretations

One paper’s abstract alone served as the focus for a class. The abstract for “Empathy and Pro-Social Behavior in Rats” (Bartal et al., 2011) outlines five individual experiments. As homework, students cartooned each experiment, all of which tested conditions under which one rat would open a transparent plastic container that restrained a second rat. Students defined the specific hypothesis being addressed in each study, the controls needed in each case (none are included in the abstract), the conclusions stated, and possible alternative interpretations.

After comparing cartoons and resolving discrepancies, the class considered whether the behaviors observed were necessarily signs of “empathy.” Might there be other explanations? Working in small groups, students proposed multiple alternatives that could in principle account for rats’ apparently helpful behavior: inquisitiveness, a pheromone signal, an aversion to squeaky distress calls, and the like. The published paper provoked substantial interest and some controversy, as reported in Nature (Gewin, 2011). We reviewed the published critique, and students found that some of “our” alternative interpretations had also been raised by top scientists in the field, again recognizing that their own thinking was scientific. Students also noted that even peer-reviewed work published in Science, where the original article appeared, can evoke intelligent criticism, and that scientists do not always agree.

Established Knowledge Can Change

A provocative set of readings discuss the discovery that peptic ulcers have a bacterial origin (Associated Press, 2005; Centers for Disease Control and Prevention, 2005). It took the PI’s ingestion of Helicobacter pylori, the suspected pathogen, hardly a canonical step in “The Scientific Method,” to generate the conclusive data. This nature of science story illustrates how established scientific knowledge—that ulcers had psychological not bacteriological etiology—can be wrong. Reading the description of Dr. Barry Marshall being met with scorn at meetings where he initially presented his unconventional hypothesis, students saw that novel (and possibly revolutionary) ideas may not be instantly welcomed. This recent scientific development highlighted the personal factors and genuine passion that can underlie science, making the point that as scientific study continues, some established ideas of today will inevitably be supplanted. The ulcer readings also illustrated the value of a healthy skepticism even about “obvious” facts, such as that the stomach’s acidity would kill all bacteria within.

Introducing Experimental Design and Peer Review

At the conclusion of many of the discussion units, the freshmen proposed follow-up experiments. The challenge: If your research team had just performed the work we reviewed, what would you do next? Each student independently devised two distinct follow-ups as homework. Three or four times during the semester, students formed teams of four to act as grant panels charged with assessing the studies designed by their peers. The first time this was done, we challenged the panels to establish appropriate funding criteria before looking at the proposed studies. Discussions of criteria led to consideration of evolution, evolutionarily conserved mechanisms, and the meaning of model systems, as many groups only wanted to fund work that is “relevant to humans.” We also discussed realities of reputation and how it may affect funding success. Some groups sought to fund “established investigators who have already published in the field,” leading other students to question how anyone gets started in research. Such discussions build students’ understanding of the sociological context of science.

After criteria had been discussed, each student submitted one of his or her experiments, sans name or other identifier, into the grant pool. The instructor then presented each proposed follow-up study to the class without evaluative comments. When the panels subsequently conferred to rank the proposed experiments, students thought critically about the work of their peers, debating and defending their judgments in the sort of open-ended give-and-take that characterizes
Using Multiple Popular Press Articles to Build Toward a Mini-Module of Primary Literature

We developed students’ critical-reading skills through repeated practice with short articles. In the process, we pointed out multiple aspects of scientific thinking, and introduced the subject matter knowledge that would be needed in the later reading of primary research reports exploring infant cognition. Early in the semester, we read and analyzed “Babies Recognize Faces Better Than Adults, Study Says” (Mayell, 2005) and a popular press account of “Plasticity of Face Processing in Infancy” (Pascalis et al., 2005), a study that tested the memories of 6- to 9-mo-old infants. Students discovered gaps in the popular press version (no information on “it” or gender distribution of infant subjects, and unclear methodology, for example). We added additional information from the Proceedings of the National Academy of Science paper as discussion required it (for details of teaching with this paper, see Hoskins, 2010b). Exercises of this sort challenge students to read actively and seek key missing information (e.g., “How many female vs. male babies were studied?” or “Exactly how was the visual training done?”) that is essential to their evaluations.

Two additional popular press stories (Talbot, 2006; Angier, 2012) and a study on babies’ perception of normal versus scrambled facial features (Maurer and Barrera, 1981) were critically analyzed in other class sessions. Discussions covered broader questions including: How can you tell whether a baby who is too young to talk notices something novel, and why might it matter? Because one of the studies was funded by the National Institutes of Health, we considered the need for eyes and other animating traits on the blocks. Other controls investigated whether babies preferred particular colors/shapes, or upward motion to downward, rather than seemingly helpful interactions (which moved the target block up) to hindering ones (which moved it down).

We started by providing the introduction, first figure, and associated text for initial analysis. As before, we did not tell the students what additional experiments were done. Through class discussion, students developed their own questions and alternative interpretations (e.g., “maybe the babies aren’t judging behavior; they just like yellow better than blue”). As in the discussions of “Babies recognize faces” and “Writing about testing . . .,” only after the students raised particular issues did we provide sections of the paper with the relevant additional information and control experiments. After analyzing the full paper, students designed follow-up experiments, vetted them in a grant panel, and then read and analyzed the authors’ actual next paper.

“Three-Month-Olds Show a Negativity Bias in Their Social Evaluations” (Hamlin et al., 2010) was concerned with younger babies’ reactions to similar social interactions. This second paper used many of the same methods as the first, facilitating students’ ability to read the material. Interestingly, the later work produced a different result, finding that younger babies were averse to hinderers but (unlike their “elders”) did not show any particular preference for helpers. As the authors discussed possible evolutionary implications of their work, we were able to return to a critical theme that had arisen earlier in the semester, in the “model systems” discussion.

Assessment in Biology 10050

The study presented here is based on tools (CAT, EDAT, SAAB) administered anonymously pre- and postcourse. To evaluate students’ understanding of course material as a basis for determining grades, we assess students in all CREATE classes using a combination of in-class activities; writing assignments; open-book, open-notes exams; and class participation. There is no assigned textbook, but students can consult during exams the notebooks/portfolios they compiled throughout the semester (see Hoskins et al., 2007, for details). We find that open-book testing changes the classroom atmosphere and relieves students from the pressure to study primarily by memorizing, making it easier for them to focus on critically evaluating scientific writing and explaining their insights. With the exception of analysis of one exam question (see Student Reactions to Emails, below), the classroom assessments were not used as data for this study.

RESULTS

CAT Outcomes

Students in the Fall CREATE Cornerstone course took the CAT (Table 1; Stein et al., 2012), and tests were scored by a trained team at Tennessee Tech University, where this test was created. Biology 10050 students’ overall CAT scores improved significantly postcourse versus precourse, with a large effect size (0.97). While there is overlap between categories, CAT questions address four main areas. Overall, the largest gains made by CREATE Cornerstone students were on CAT questions that tested “evaluating and interpreting information.” Students also made gains on questions involving problem solving, creative thinking, and/or effective communication (the other three subcategories addressed by the CAT). While these findings must be interpreted with caution due to the small sample size, they suggest that students in the pilot CREATE Cornerstone course made substantial gains in their
ability to read, understand, and critically analyze information, and that such gains are transferable to the content domain addressed by the CAT test, which was not related to the material covered in the course.

**EDAT Outcomes**

Students in both Fall and Spring CREATE Cornerstone classes completed a pre- and postcourse EDAT that was scored using a 10-point rubric (Sirum and Humburg, 2011). Results are summarized in Table 2. Scores suggest that the first-year students gained significantly in experimental design ability over the semester, citing more components of an “ideal” experimental design postcourse than precourse.

**SAAB Outcomes**

Results from the SAAB surveys for each class are displayed in Table 3. The upper group reflects the items related to students’ epistemological beliefs about science (see Hoskins et al., 2011, for a discussion of the derivations of all categories).

SAAB results show significant gains made by CREATE Cornerstone students in all six skills and attitudes categories and in the majority (four out of seven) of epistemological categories. Students in the upper-level CREATE course (for which a year of introductory biology, a semester of genetics, and a semester of cell biology are prerequisites) shifted significantly on all skills and attitudes categories, and three of the seven epistemological categories. Students in the mid-level physiology course (for which a year of introductory biology and a semester of genetics are prerequisites), in contrast, did not shift significantly in any category.

Effect sizes help to determine whether statistically significant changes are likely to be meaningful. For skills and attitudes shifts, effect sizes for freshmen were large (Cohen, 1992) in five of the six categories and moderate for “interpreting data.” Effect sizes for upper-level CREATE students in these categories were all large. In this regard, it may be relevant that upper-level students read literature that was substantially more complex and looked closely at more figures during the semester than did the first-year students. It is also interesting to note that the mid-level physiology course included a weekly laboratory, in which data were generated and analyzed, and one experimental design activity.

For epistemological beliefs categories, effect sizes in three of the four categories that shifted significantly in the freshman CREATE group (certainty of knowledge, innate ability, creativity of science) were moderate. The effect size of “sense of scientists as people” was large. Upper-level CREATE students also shifted significantly in this category, but with a smaller effect size, possibly reflecting the fact that many upper-level students were working in labs and had a better sense precourse of what research scientists were like. Upper-level CREATE students also showed significant changes in understanding of the uncertainty of scientific knowledge (large effect size), and of “sense of scientists’ motivations” (moderate effect size).

Both the CREATE courses, but not the mid-level physiology course, sent email surveys to authors of papers and discussed author responses late in the semester. Different material was read and analyzed in each CREATE course; thus, different authors were queried and different responses were received by the two groups. We think it likely that this component of the CREATE courses played a large role in changing students’ opinions about what scientists are like and (for upper-level CREATE students) why they do what they do.

**Student Reactions to Emails**

On the second exam in the Spring semester, we included a question asking students about their reactions to the author emails, focusing on their preconceptions about “scientists/research careers” and whether the author responses changed these views. We coded all the responses (n = 15), extracting key themes from each, and summarize below the themes mentioned by four or more students.

The most prevalent response to the emails was students’ statements that, precourse, they had assumed today’s researchers were “straight-A” students in college (14/15 responses; 93% of students). The same students (14/15) noted that they no longer believed this to be true, citing several authors who described academic struggles that preceded their eventual success. Thirteen out of 15 students (86%) said that the responses had changed their preconceptions about researchers, and 9/15 (60%) noted that respondents stressed the importance of passion (as opposed to good grades) as a key to research success. Seven out of 15 students (47%) expressed enthusiasm on learning that the responding scientists described a great deal of work-related travel, including international travel. Forty percent of students (6/15) described having held one or more of the preconceptions that 1) scientists were loners or nerds, 2) who lacked social lives, 3) because science consumed all their time. A similar

### Table 1. CAT test results

<table>
<thead>
<tr>
<th>Critical Thinking Ability Test (CAT)</th>
<th>Precourse</th>
<th>Postcourse</th>
<th>n</th>
<th>Significance</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>9.6 (2.5)</td>
<td>13.0 (4.4)</td>
<td>15</td>
<td>p &lt; 0.05</td>
<td>0.97</td>
</tr>
</tbody>
</table>

*The CAT (duration 1 h) was administered pre- and postcourse to the Fall 2011 Biology 10050 class and scored at Tennessee Tech University. We present the overall score for the test, precourse vs. postcourse. Fifteen students took both tests. Significance: Student’s t test.

### Table 2. EDAT results: mean and SD

<table>
<thead>
<tr>
<th>EDAT test</th>
<th>Precourse</th>
<th>Postcourse</th>
<th>n</th>
<th>Significance</th>
<th>Effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>4.3 (2.1)</td>
<td>5.9 (1.4)</td>
<td>28</td>
<td>p &lt; 0.01</td>
<td>0.91</td>
</tr>
</tbody>
</table>

*Pool of two classes of Biology 10050; n = 28 total. Statistical significance tested with Wilcoxon signed-rank test. Scores can range from 0 to 10, per the EDAT rubric (see Sirum and Humburg, 2011).
Table 3. SAAB survey outcomes in three student cohorts: freshman CREATE students \((n = 28)\), upper-level CREATE students \((n = 19)\), and mid-level non-CREATE students \((n = 23)\)

<table>
<thead>
<tr>
<th>Category</th>
<th>Precourse mean</th>
<th>Postcourse mean</th>
<th>Significance</th>
<th>Effect</th>
<th>#Ss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decoding literature</td>
<td>17.3 (3.2)</td>
<td>21.9 (3.0)</td>
<td>&lt;0.001</td>
<td>1.48</td>
<td>6</td>
</tr>
<tr>
<td>Interpreting data</td>
<td>14.1 (2.6)</td>
<td>15.4 (2.3)</td>
<td>0.008</td>
<td>0.53</td>
<td>4</td>
</tr>
<tr>
<td>Active reading</td>
<td>13.0 (2.4)</td>
<td>16.1 (2.3)</td>
<td>&lt;0.001</td>
<td>1.32</td>
<td>4</td>
</tr>
<tr>
<td>Visualization</td>
<td>12.5 (2.8)</td>
<td>15.6 (2.1)</td>
<td>&lt;0.001</td>
<td>1.27</td>
<td>4</td>
</tr>
<tr>
<td>Think like a scientist</td>
<td>11.9 (2.2)</td>
<td>15.5 (1.6)</td>
<td>&lt;0.001</td>
<td>1.90</td>
<td>4</td>
</tr>
<tr>
<td>Research in context</td>
<td>10.9 (1.5)</td>
<td>13.8 (1.4)</td>
<td>&lt;0.001</td>
<td>2.00</td>
<td>3</td>
</tr>
<tr>
<td>Certainty of knowledge</td>
<td>22.1 (2.9)</td>
<td>24.3 (3.8)</td>
<td>0.002</td>
<td>0.66</td>
<td>6</td>
</tr>
<tr>
<td>Ability is innate</td>
<td>6.9 (1.4)</td>
<td>8.0 (1.6)</td>
<td>0.005</td>
<td>0.73</td>
<td>2</td>
</tr>
<tr>
<td>Science is creative</td>
<td>3.9 (0.8)</td>
<td>4.3 (0.7)</td>
<td>0.005</td>
<td>0.53</td>
<td>1</td>
</tr>
<tr>
<td>Scientists as people</td>
<td>2.8 (0.9)</td>
<td>3.6 (1.0)</td>
<td>0.004</td>
<td>0.84</td>
<td>1</td>
</tr>
<tr>
<td>Scientists’ motives</td>
<td>3.8 (1.1)</td>
<td>4.1 (0.8)</td>
<td>ns</td>
<td>0.35</td>
<td>1</td>
</tr>
<tr>
<td>Known outcomes</td>
<td>3.9 (1.0)</td>
<td>4.2 (0.9)</td>
<td>ns</td>
<td>0.32</td>
<td>1</td>
</tr>
<tr>
<td>Collaboration</td>
<td>4.2 (0.66)</td>
<td>4.3 (0.8)</td>
<td>ns</td>
<td>0.14</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Precourse mean</th>
<th>Postcourse mean</th>
<th>Significance</th>
<th>Effect</th>
<th>#Ss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decoding literature</td>
<td>15.5 (2.8)</td>
<td>20.3 (2.7)</td>
<td>&lt;0.001</td>
<td>1.75</td>
<td>6</td>
</tr>
<tr>
<td>Interpreting data</td>
<td>13.7 (2.2)</td>
<td>16.4 (1.7)</td>
<td>&lt;0.001</td>
<td>1.39</td>
<td>4</td>
</tr>
<tr>
<td>Active reading</td>
<td>13.9 (2.1)</td>
<td>16.9 (1.9)</td>
<td>&lt;0.001</td>
<td>1.50</td>
<td>4</td>
</tr>
<tr>
<td>Visualization</td>
<td>13.3 (2.1)</td>
<td>16.6 (1.7)</td>
<td>&lt;0.001</td>
<td>1.74</td>
<td>4</td>
</tr>
<tr>
<td>Think like a scientist</td>
<td>13.3 (2.5)</td>
<td>16.5 (2.1)</td>
<td>&lt;0.001</td>
<td>1.39</td>
<td>4</td>
</tr>
<tr>
<td>Research in context</td>
<td>13.5 (1.1)</td>
<td>14.3 (0.9)</td>
<td>0.037</td>
<td>0.80</td>
<td>3</td>
</tr>
<tr>
<td>Certainty of knowledge</td>
<td>23.0 (2.7)</td>
<td>26.1 (2.9)</td>
<td>0.021</td>
<td>0.82</td>
<td>6</td>
</tr>
<tr>
<td>Ability is innate</td>
<td>7.3 (1.9)</td>
<td>8.4 (1.5)</td>
<td>ns</td>
<td>0.65</td>
<td>2</td>
</tr>
<tr>
<td>Science is creative</td>
<td>4.1 (0.7)</td>
<td>4.6 (0.9)</td>
<td>ns</td>
<td>0.63</td>
<td>1</td>
</tr>
<tr>
<td>Scientists as people</td>
<td>2.5 (1.0)</td>
<td>3.9 (0.8)</td>
<td>0.004</td>
<td>0.44</td>
<td>1</td>
</tr>
<tr>
<td>Scientists’ motives</td>
<td>3.8 (1.0)</td>
<td>4.3 (0.6)</td>
<td>0.014</td>
<td>0.63</td>
<td>1</td>
</tr>
<tr>
<td>Known outcomes</td>
<td>3.7 (1.0)</td>
<td>4.3 (1.1)</td>
<td>ns</td>
<td>0.57</td>
<td>1</td>
</tr>
<tr>
<td>collaboration</td>
<td>4.4 (0.6)</td>
<td>4.5 (0.6)</td>
<td>ns</td>
<td>0.17</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>Precourse mean</th>
<th>Postcourse mean</th>
<th>Significance</th>
<th>Effect</th>
<th>#Ss</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decoding literature</td>
<td>19.6 (4.3)</td>
<td>20.4 (3.3)</td>
<td>ns</td>
<td>0.21</td>
<td>6</td>
</tr>
<tr>
<td>Interpreting data</td>
<td>15.3 (2.3)</td>
<td>15.7 (2.6)</td>
<td>ns</td>
<td>0.16</td>
<td>4</td>
</tr>
<tr>
<td>Active reading</td>
<td>14.7 (2.3)</td>
<td>14.7 (2.4)</td>
<td>ns</td>
<td>0.01</td>
<td>4</td>
</tr>
<tr>
<td>Visualization</td>
<td>14.0 (2.3)</td>
<td>14.7 (1.8)</td>
<td>ns</td>
<td>0.34</td>
<td>4</td>
</tr>
<tr>
<td>Think like a scientist</td>
<td>13.8 (2.7)</td>
<td>13.8 (2.5)</td>
<td>ns</td>
<td>0.00</td>
<td>4</td>
</tr>
<tr>
<td>Research in context</td>
<td>13.5 (1.1)</td>
<td>13.5 (1.4)</td>
<td>ns</td>
<td>0.40</td>
<td>3</td>
</tr>
<tr>
<td>Certainty of knowledge</td>
<td>23.7 (3.3)</td>
<td>23.6 (3.2)</td>
<td>ns</td>
<td>–0.03</td>
<td>6</td>
</tr>
<tr>
<td>Ability is innate</td>
<td>7.3 (1.5)</td>
<td>7.6 (1.5)</td>
<td>ns</td>
<td>0.20</td>
<td>2</td>
</tr>
<tr>
<td>Science is creative</td>
<td>3.8 (1.2)</td>
<td>4.2 (0.7)</td>
<td>ns</td>
<td>0.42</td>
<td>1</td>
</tr>
<tr>
<td>Scientists as people</td>
<td>3.1 (0.9)</td>
<td>3.3 (1.1)</td>
<td>ns</td>
<td>0.03</td>
<td>1</td>
</tr>
<tr>
<td>Scientists’ motives</td>
<td>3.9 (0.9)</td>
<td>4.0 (1.1)</td>
<td>ns</td>
<td>0.10</td>
<td>1</td>
</tr>
<tr>
<td>Known outcomes</td>
<td>3.9 (1.0)</td>
<td>4.0 (0.8)</td>
<td>ns</td>
<td>0.11</td>
<td>1</td>
</tr>
<tr>
<td>Collaboration</td>
<td>4.4 (0.5)</td>
<td>4.2 (0.6)</td>
<td>ns</td>
<td>–0.36</td>
<td>1</td>
</tr>
</tbody>
</table>

*a*Responses were tabulated using a 1–5 scale. (1 = “I strongly disagree”; 2 = “I disagree”; 3 = “I am neutral”; 4 = “I agree”; 5 = “I strongly agree”). Some propositions were worded so that an answer reflecting a more mature understanding would get a lower score (“I accept the information about science presented in newspaper articles without challenging it,” for example). These were reverse-scored for analysis.

The Wilcoxon signed-rank test for statistical significance was performed on precourse/postcourse raw data totals for all categories. Category 1–6: self-rated skills and attitude factors; categories 7–13: epistemological factors.

The survey was developed in a previous study of upper-level CREATE students (Hoskins *et al*., 2011). Different categories are probed by different numbers of statements (#Ss).

*b*p values for statistical significance (Wilcoxon signed-rank test). ns = not significant.

*c*Mean difference/average SD.

*d*#Ss = number of statements in category.
percentage noted that precourse they had assumed all scientists had lofty goals of “helping people,” but they had come to realize that many had more personal goals of satisfying their own curiosity. Five out of 15 students (33%) stated that precourse they had assumed most scientists did not enjoy their jobs, that research was not fun, and that lab life was boring, but they no longer held these views. Five out of 15 (33%) said they were surprised to learn scientists had flexible work schedules, and a similar percentage stated that they had learned from the emails that motivation was very important. Finally, 4/15 (27%) noted their surprise that the authors

**DISCUSSION**

**Genesis of the CREATE Strategy**

The CREATE strategy originated as a response to the observation that many upper-level undergraduate biology majors—despite the years spent studying a wide range of scientific topics—were not well-prepared to read and understand primary literature; did not readily “think like scientists,” with an appropriately critical eye; did not see science research as an attractive career choice; and had little or no practical experience mustering their content knowledge to attack novel scientific problems. Discussions with students in other courses over the years, and with other faculty on our campus and elsewhere, revealed that many students believed: research is dull, and lab exercises formulaic and boring (Luckie et al., 2004); there is a single and eternal right answer to every scientific question (Li and Tsai, 2008); primary literature is written in a nearly unbreakable code; and scientists themselves are stereotypic nerds or “machinery kind of people” (Hoskins et al., 2007). Our findings in the pilot CREATE Cornerstone course suggest that these viewpoints can be changed over a single semester through intensive analysis of scientific literature.

**Themes Highlighted in Readings**

The curriculum examples outlined above illustrate how fundamental features of scientific thinking can be studied in a realistic domain-specific context, which appears to be a key element in developing critical-thinking skills (Willingham, 2007). Students repeatedly thought carefully about control groups—what they “control” for, how they are interpreted, and why they are needed. Multiple studies underscored the importance of careful attention to sample size and selection. In the experiments on infants, for example, students raised issues of possible gender-related behavioral differences, whether postnatal age is comparable between full-term and premature infants, and the like. Students practiced developing alternative interpretations of data and noted that not all conclusions are equally strong. Several studies highlighted the potential for introducing unanticipated bias (see discussion of a possible “Clever Hans” effect in “Babies Recognize Faces Better Than Adults, Study Says” in Hoskins, 2010b). Students saw that original, interesting, and important investigations are currently ongoing (many readings were published in 2011–2012). Students also recognized that even very early in their academic careers they are capable of reading, understanding, and intelligently criticizing scientific literature, and that research science is neither routine, predictable, or boring, nor something found only in textbooks.

**Grant Panels Promote Open-Ended Thinking and Insight into the Nature of Science.** CREATE Cornerstone students made significant gains on the EDAT, which presents a scenario distinct from that of any of the Cornerstone readings. Students’ gains on this test suggest that their general experimental design skills have improved during the semester.

Experimental design skills are honed in class through grant panel activities that focus on follow-up experiments to the studies we analyzed that are designed by the students as homework. These are repeated several times during the semester. Although panels focus specifically on experimental systems under study in class, they likely help students develop a more generalized skill in experimental design and creative thinking. In each panel all students’ experiments are reviewed, and the panels (groups of four students) discuss the merits of each. Early in the semester, some experiments must be culled based on absence of a hypothesis, absence of a cartoon, or general lack of clarity (approximately five of 20 in early panels). In end-of-semester exercises, virtually every experiment meets the basic criteria and can be considered seriously. Statements of hypotheses become clearer, controls stronger, designs and procedures better illustrated, and potential outcomes well anticipated.

Besides likely contributing to the development of students’ experimental design skills, the grant panels provide insights into the nature of science. It becomes evident, as the activity is repeated during the semester that, among the top experiments (typically four or five stand out), the study perceived by a particular panel to be “best” is to some degree a matter of taste. Some students prefer a reductionist approach, others an expansion of the study to encompass additional sensory modalities (e.g., an experiment investigating whether babies learn to recognize faces faster if each face is associated with a different musical tune). Some students focus mainly on experiments aimed at developing treatments for humans (e.g., take genes involved in planar regeneration and immediately seek their counterparts in mammals). Many of our students are accustomed to “textbook” science where, typically, only the (successful) end points of studies are described, and very little current-day work is featured. The grant panel activity introduces the idea that working scientists likely select their follow-up experiment from a variety of valid possibilities, and that personal styles and preferences could influence such decisions.

**Critical-Thinking and Experimental Design Skills—Tools of Science.** A significant number of students show interest in science in high school or before (often significantly before [Gopnik, 2012]), but do not pursue STEM studies at the tertiary level. Either they never consider studying science in college, or they switch out of the field for a variety of reasons in their first or second year (Seymour and Hewitt, 1997; Committee on Science and Technology, 2006). At the same time, for students who persist in STEM majors, some of the most creatively challenging and thought-provoking courses—capstone experiences—are reserved for seniors (Goyette and DeLuca, 2007; Usher et al., 2011; Wiegant et al., 2011). We hoped to convey some of the analytical and creative aspects of science at the outset of students’ college careers with a CREATE course designed for
freshmen. Providing this training early in students’ academic experience might help students gain skills and develop attitudes that would support their persistence in STEM (Harrison et al., 2011).

We used the CAT and EDAT assessments to probe the development of students’ abilities as they practiced the literature analysis process. The CAT test focuses on a content domain distinct from that of the CREATE class but challenges students in some parallel ways. Students must determine what data mean, decide which data are relevant, draw conclusions based on their understanding, and explain themselves in writing. Many campuses are using the CAT test for programmatic assessment, comparing scores of freshman with those of seniors, for example. We are aware of only one published study using CAT in a pre/post, single-course situation. First-year students in a semester-long inquiry-based microbiology module at Purdue University, performing hands-on research in an introductory class, make significant CAT gains during the semester (Gasper et al., 2012). The finding that CREATE Cornerstone students at CCNY similarly made significant gains on this test in a single semester suggests that transferable critical-thinking skills, such as those measured by the CAT, can also be built through classroom activities that do not involve hands-on inquiry labs.

While the small sample size in this pilot study precludes broad conclusions, it is interesting that our students made the largest gains on CAT questions whose solution required “evaluation and interpretation.” Introduction to Scientific Thinking emphasizes looking closely at data, reconstructing the experiment or study that gave rise to the data, and reasoning carefully about the logic of interpretations and the significance of the findings. Students carry out this process in a variety of content domains, engaging in friendly arguments about whether rats are empathic or just noise-averse, whether writing about fears really prevents chewing on tests, and what it is that babies might prefer about a yellow square with googly eyes (the color? the shape? the eyes? the “helpful” behavior?). As noted by Stanger-Hall (2012), close to 80% of U.S. high school seniors performed below the science proficiency level on a recent national standardized test (National Center for Education Statistics, 2009). Among undergraduates, barely more than half the students sampled at 24 institutions made gains in critical thinking during their first 2 yr of college, as measured by the Collegiate Learning Assessment (Arum and Roksa, 2011). These data suggest that current course work in high school and during early college years (when standard introductory science courses are taken by STEM majors) is not promoting substantial development of higher-order thinking and analytical reasoning skills. We find CREATE Cornerstone students’ outcomes on the CAT assessment encouraging in this regard. At the same time, some researchers suggest results of low-stakes tests like the Collegiate Assessment of Academic Proficiency may be influenced by low performance motivation among test takers, because participation in such exercises has no bearing on class grade (Wise and DeMars, 2005). This issue could potentially influence our students’ performance on anonymous assessments. While we have no independent measure of students’ motivation for participating in our study, we believe it likely that, as participants in a novel course, they find the opportunity to be part of a scientific study to be intriguing and a motive to perform well.

The EDAT assessment called on students to think like scientists: analyze a problem, determine evidence required to solve it, and design a properly controlled experiment that could generate the relevant data. Students made statistically significant gains in their experimental design ability, with their postcourse responses mentioning more of the points that experts see as essential to good experimental design (Sirum and Humburg, 2011). In the Cornerstone classroom, students repeatedly proposed and evaluated experimental designs as they participated in multiple grant panels and worked with different student-colleagues. We suspect that these exercises served as a form of practice during the CREATE semester, helping students build competence in their ability to formulate, express, and defend ideas about particular proposed studies (Ambrose et al., 2010). At the same time, the challenge of producing an experiment that would be singled out by a grant panel for “funding” may have stimulated some students’ efforts to be particularly creative in their experimental designs.

The CAT and EDAT findings also support our sense that skills deemed important by many science faculty (e.g., problem solving/critical thinking, data interpretation, written and oral communication; Coil et al., 2010), including ourselves, can be taught in a course that emphasizes the process of science, including close reading and critical analysis of primary literature, creative experimental design, and a look behind the scenes into the lives and dispositions of paper authors. While we teach or review relevant content in the context of particular reading assignments, we do not seek to achieve the broad coverage of a typical introductory course. Students need not know the details of the electron transport chain in order to analyze “Babies Recognize Faces Better Than Adults, Study Says,” although they do need to know the fundamental logic of study design, use of controls, and the danger of being unwilling to think beyond your preferred hypothesis. To analyze the rat studies, students must understand the terms “empathy” and “prosocial behavior,” and know how to think about variables, controls, and multiple aspects of animal behavior. In each case, they also need metacognitive awareness—the ability to determine what they do and do not understand, as well as “how we know what we know” (Tanner, 2012), another skill developed through practice during the semester.

Student Attitudes and Beliefs—Influences on Learning and Career Options. On the SAAB survey, freshmen reported significant gains in their self-rated ability to: “decode” primary literature; interpret data; read actively (annotating, concept mapping and/or cartooning the material they were reading); visualize scientific procedures; feel like they were thinking like scientists; and see experiments in a broader context (Table 3). Effect sizes (Cohen, 1992; Cote, 2002) were large for five of the SAAB measures and moderate for “interpreting data.” With regard to students’ epistemological beliefs, previous researchers (Perry, 1970; Baxter Magolda, 1992) have noted that students’ naïve epistemological beliefs about science resist change, even after a 4-yr undergraduate program. In some cases, such beliefs appear to regress after students take introductory biology courses (Smith and Wenk, 2006; Samsar et al., 2011). After one semester, the freshmen in the CREATE Cornerstone course reported significant increases in four of the seven epistemological categories we surveyed:
the uncertain nature of scientific knowledge; the question of whether one needs to have a special innate ability to do science; whether science is creative; and their sense of scientists as “real people.” A concurrent upper-level CREATE class also made gains in several epistemological categories, while students in a non-CREATE comparison course that included a weekly laboratory session did not change significantly in any category (Table 3). These findings argue that the shifts we see relate to the CREATE experience, rather than to intellectual maturation that might occur naturally in college biology students over the course of a semester.

While student epistemology is rarely emphasized in college teaching handbooks, students’ attitudes in this area can strongly influence their learning. For example, students who feel that intelligence is a fixed quantity in which they are lacking may decrease their efforts to learn and study ineffectively as a result (Henderson and Dweck, 1990). The high attrition rate of students from the biology major has been attributed in large part to students’ failure to connect intellectually with the subject, and the traditional mode of teaching introductory courses itself can slow students’ development of higher-order thinking skills (e.g., analysis, synthesis, evaluation; Bloom et al., 1956). While the majority of faculty members who teach introductory biology courses want students to learn higher-order skills, exams in such courses tend to focus at lower levels (Momsen et al., 2010). Multiple-choice testing (often considered a practical requirement for a large lecture course) shapes students’ study habits in unproductive ways and interferes with critical thinking (Stanger-Hall, 2012). Noting that epistemological change is typically slow, Smith and Wenk point out that “…one cannot ignore the potential retarding effect of an entrenched instructional system of lecture, textbook readings, and recitation on the students’ epistemological development” (Smith and Wenk, 2006, p. 777). This phenomenon may be reflected in the differences in responses of CREATE and non-CREATE students on the SAAB survey.

The change in first-year students’ attitudes about scientists as people is large. We saw previously that upper-level students harbored negative opinions about scientists and the research life (Hoskins et al., 2007), but we did not know whether these ideas developed during college or before. Finding that first-year students also assumed, pre-course, that scientists were antisocial and that research careers were dull suggests that students finish high school and enter college with negative conceptions about research/researchers. Multiple years of college science education apparently do little to change these ideas. The shift we saw in students’ views of scientists and research careers is likely attributable to one of the more unconventional techniques used in CREATE classes, the email survey of paper authors (see analysis of email reactions in Results, above). This student response to a Cornerstone exam prompt regarding author replies is typical:

I had this preconception [pre-course] that…you had to be like Einstein to penetrate that field. I thought you always had to have straight A’s and be highly versatile, but after reading the e-mails from the authors I know that’s definitely not the case. From what they said I know that you don’t have to be perfect or like Einstein. It’s the passion and motivation to learn and make discoveries. You have to have a drive that leads you on. It was inspiring to hear that “science has many paths” by S—[the quote in the author’s response was “there are many paths to science”]. To me this means that there’s no one path or just one requirement like reading a textbook, but many. I can conduct research with as little a space as a backyard or in one of the biggest labs and each one could lead to success and greatness. (Exam response, freshman in Biology 10050)

CREATE Freshmen—Thinking/Attitude Gains

Students’ self-reported reactions to author emails suggest that students starting college, at least at CCNY, harbor serious misconceptions about research/researchers that could likely interfere with their potential development as scientists. Nearly all students in the class noted that before they read the authors’ responses they had assumed that “only straight A students can become scientists.” This supposition changed when responding scientists recounted particular academic travails (e.g., rejection from some graduate schools) that preceded their success. Other student comments regarding their prerequisite suppositions that research is boring; that researchers are both overworked and unhappy with their jobs; and that such jobs allow no time for hobbies, families, or personal life, suggest that students’ precollege science experience has not presented research careers in an accurate light. Notably these views defy logic, suggesting that some adopted the stereotype without giving it much thought. Why would people who are so smart (“like Einstein”) and who achieved “straight A” in college choose dull, boring careers? Why would someone engaged in a boring career that he or she did not enjoy, nevertheless work so intensely that he or she had time for nothing else? We have speculated elsewhere that popular culture’s depictions of scientists may influence students negatively, starting in high school or before (Hoskins and Stevens, 2009). Changing students’ negative views of researchers/research careers is a likely required first step, if such students are to be inspired to undertake undergraduate research experiences that can lead to research careers (Harison et al., 2011). Given that the no-cost email survey of authors can have a strong positive impact on students’ views, we encourage other STEM faculty, particularly those working with high school students or first-year undergraduates, to consider this activity.

Early Interventions. Traditionally, STEM-inclined students spend their early college years in conventional core courses. Electives, including capstone courses, are reserved for upper-level students. Recently, however, a number of colleges and universities have begun developing nontraditional courses for entering students. A 5-d presemester “boot camp” for biology students at Louisiana State University aims to teach students about the different expectations at college versus high school, focusing on study strategies and introductory biology material. This brief presemester experience resulted in gains for boot-camp veterans, as compared with a matched control group, in classroom performance and in persistence in the major (Wischusen and Wischusen, 2007). In a new course focused on freshmen’s ability to reason scientifically, students studied a variety of topics that a faculty group had deemed valuable for introductory STEM courses. Students made significant gains in understanding of control of variables and proportional thinking, and also showed greater persistence in STEM (Koenig et al., 2012). Freshmen at Cabrini College participated in Phage Genomics rather than a standard Introductory Biology laboratory course. The novel course involved participation in a two-semester, hands-on research project
that significantly increased students’ interest in postgraduate education, their understanding of scientific research, and their persistence in the biology major (Harrison et al., 2011).

### Beyond Textbooks

Visual representations in journal articles are both more frequent and more complex than those seen in textbooks (Rybarczyk, 2011). When visual representations do appear in textbooks, they rarely illustrate the process of science (Duncan et al., 2011). Controversy, certainly a part of science, is virtually absent from textbooks (Seethaler, 2005). Some faculty members feel that encounters with primary literature, as well as capstone courses, and the majority of undergraduate research experiences, should be reserved for upper-level students, who have built a broad foundation of content knowledge in textbook-based courses. We agree that understanding the nuts and bolts of a paper is a prerequisite for full understanding. Further, analysis and comprehension skills are better taught in the context of a particular content domain (Willingham, 2007). At the same time, particularly for biology, the explosion of fundamental content makes it impossible for faculty to cover, let alone teach, “basic” material even to the same depth it was covered in the introductory courses of their own undergraduate years (Hoskins and Stevens, 2009). In addition, despite having encountered them in multiple courses, students may fail to retain key concepts (e.g., function of control experiments; see Shi et al., 2011). Our compromise was to base a freshman course on particular examples of scientific literature, choosing topics in a limited range of content areas and focusing in-depth on scientific thinking and data analysis. While we designed Introduction to Scientific Thinking for future biology majors, the approach could be easily adapted to other STEM domains. Interestingly, a recent study argues that long-term cognitive advantages can arise from studying individual topics in depth. First-year undergraduates’ grades in science courses were highest for students who had studied a single topic in depth for a month or more at any time during high school. Grades showed no correlation with economic status, region, class size, parents’ academic level, or other factors (Schwartz et al., 2009).

Taken together, our findings support the hypothesis that a CREATE Cornerstone course designed for first-year students can bring about gains in multiple areas, including critical-thinking and experimental design ability, self-rated attitudes, abilities and epistemological beliefs, and understanding of scientists as people. Our freshman students did not have a base of content knowledge in biology beyond what they retained from high school or had absorbed from popular media. By choosing articles from top journals (e.g., Science, Nature) but focusing on topics that did not require deep understanding of, for example, gene knockout techniques or electrophoresis, we were able to give students a taste of the sorts of design logic, interpretational challenges and controversies, and creativity that are hallmarks of real-world scientific investigation. At the same time that our students gained understanding of how authentic scientific studies are carried out and interpreted, their email interviews of authors provided a personalized glimpse behind the scenes into the lives, attitudes, and motivations of the researchers themselves. Ideally, such insights will help to dispel misconceptions that can drive students away from science. To the extent that students in the one-semester Introduction to Scientific Thinking course make significant gains in scientific thinking ability, they become better prepared to master the material in any STEM major they choose, as gains in critical-thinking and reading/analytical skills should help them manage the information load in the more content-heavy science courses to come.

### CONCLUSIONS

Introduction to Scientific Thinking, the CREATE Cornerstone course, improved critical-thinking and experimental design skills of freshmen at the same time that it positively shifted their attitudes about their reading/analytical abilities, their understanding of scientists as people, and multiple aspects of their epistemological beliefs. There are few reported approaches to changing the critical thinking of first-year science students, and it appears that epistemological beliefs among college students at all undergraduate levels are quite stable. We find that a one-semester course positively affects both. The course has no laboratory component, so it is relatively inexpensive to offer. Because the topic area of the articles that can be analyzed ranges broadly, readings can be selected for their utility in a variety of introductory science courses. Finally, the email survey responses from paper authors have a strong effect on students’ sense of scientists as people, helping them to overcome misconceptions of the sort that can dissuade students from seeking research opportunities and, by extension, research careers. We are encouraged by the results of this pilot study and conclude that important gains—both practical and attitudinal—with potential to help students make progress in STEM, can be achieved in a one-semester course that meets 2.5 h/wk and could, in principle, be added to many curricula.

If, as exhorted by many science-education policy reformers, we are to do a better job at encouraging students to consider research careers seriously (National Research Council, 2003; American Association for the Advancement of Science, 2011), we need to move beyond standard first-year courses and reveal scientific research as a creative and exciting career choice undertaken by interesting and diverse individuals, not unlike the first-year students themselves. While it would be gratifying to see more students enter STEM research fields, the enhancement of skills, attitudes, and epistemological beliefs concerning science engendered by CREATE Cornerstone is aligned with societal and civic goals, even for students who go in other directions.

### ACKNOWLEDGMENTS

Many thanks to CCNY students for participating in the CREATE courses and/or associated assessments. We thank Dr. Millie Roth and NKem Stanley-Mbamelu, Constance Harper, and Maria Harvey of City College Academy for Professional Preparation for assistance with first-year students, and Drs. Shubha Govind, Anu Janakiram, Kristy Kenyon, Jonathan Levitt, and Leslie Stevens for assistance, comments on the manuscript, and ongoing discussions of CREATE teaching/learning issues. Many thanks to Dr. Barry Stein, Elizabeth Lisc, and the CAT team at Tennessee Tech University for use and grading of the CAT instrument. We also thank the anonymous reviewers, whose insightful comments strengthened the manuscript. We are very grateful to the NSF for support.

### REFERENCES


Hoskins SG (2010a). “But if it’s in the newspaper, doesn’t that mean it’s true?”: developing critical reading & analysis skills by evaluating newspaper science with CREATE. Am Biol Teach 72, 415–420.


Poldrack R (2011b). NYT letter to the editor: The uncut version. www.russpoldrack.org/search?updated-min=2011-01-01T00:00:00-08:00&updated-max =2012-01-01T00:00:00-08:00&max_results =8 (accessed 11 September 2012).


Mutation-Based Learning to Improve Student Autonomy and Scientific Inquiry Skills in a Large Genetics Laboratory Course

Jinlu Wu

Department of Biological Sciences, Faculty of Life Science, National University of Singapore, Singapore 117543

Monitoring Editor: Diane K. O'Dowd

Laboratory education can play a vital role in developing a learner's autonomy and scientific inquiry skills. In an innovative, mutation-based learning (MBL) approach, students were instructed to redesign a teacher-designed standard experimental protocol by a “mutation” method in a molecular genetics laboratory course. Students could choose to delete, add, reverse, or replace certain steps of the standard protocol to explore questions of interest to them in a given experimental scenario. They wrote experimental proposals to address their rationales and hypotheses for the “mutations”; conducted experiments in parallel, according to both standard and mutated protocols; and then compared and analyzed results to write individual lab reports. Various autonomy-supportive measures were provided in the entire experimental process. Analyses of student work and feedback suggest that students using the MBL approach 1) spend more time discussing experiments, 2) use more scientific inquiry skills, and 3) find the increased autonomy afforded by MBL more enjoyable than do students following regimented instructions in a conventional “cookbook”-style laboratory. Furthermore, the MBL approach does not incur an obvious increase in labor and financial costs, which makes it feasible for easy adaptation and implementation in a large class.

INTRODUCTION

Laboratory education plays various roles in developing students’ interests, scientific inquiry skills, and understanding and application of scientific concepts learned in lecture. It is believed that appropriate exposure to practical work is an essential component of any bioscience degree (Hofstein and Lunetta, 2004; Hofstein and Mamlok-Naaman, 2007). However, laboratory work is expensive; once a laboratory has been set up, it leaves little room for changes, due to budget and labor constraints. A direct consequence of this limitation is a “cookbook” approach widely used in laboratory teaching. In this approach, students often simply go through the motions of laboratory work in a series of cookbook-style activities. Furthermore, experiments are repeated for many batches of students and are guaranteed to produce expected results. The very stringent protocol and highly predictable results reduce student autonomy, curiosity, and motivation. Thus, the effectiveness of this conventional cookbook-style laboratory education has often been questioned (White, 1996; Adams, 2009).

It is widely accepted that giving learners autonomy increases motivation (Sierens et al., 2009). According to the self-determination theory, when the three basic psychological needs of autonomy, competence, and relatedness are met, intrinsic motivation is enhanced, leading to autonomous internalization of behaviors of initial extrinsic origin (Ryan and Deci, 2000; Wichmann, 2011). The needs for autonomy, competence, and relatedness refer to the experience of behavior as volitional and reflectively self-endorsed, effectively enacted, and intimately connected to others, respectively (Katz and Assor, 2007; Niemiec and Ryan, 2009; Van den Broeck et al., 2010). The need for autonomy is particularly important in promoting intrinsic motivation (Ryan and Deci, 2000). Autonomy-supportive teaching has been demonstrated to
enhance student intrinsic motivation and ownership and to promote a more enduring psychological investment in deep-level thinking (Sheldon, 1995; Chan, 2001; Stefanou et al., 2004; Katz and Assor, 2007; Furtak and Kunter, 2012). Autonomy support can be manifested in at least three different ways (Stefanou et al., 2004): organizational (developing rules, latitude over rate of progress, selecting due dates for assignments), procedural (choosing the appropriate media to present ideas and results), and cognitive autonomy supports (justifying an argument, generating one’s own solutions, evaluating ideas and results). The cognitive autonomy support is the most critical, leading to psychological investment in learning. However, in many traditional laboratory exercises, no deviation is allowed, and no choice is offered to support student autonomy in the design and performance of experiments.

To achieve efficient laboratory teaching and learning, extensive exploration of reformed pedagogical approaches has been undertaken. The project-based laboratory was used to develop initiative and innovation (Cioc et al., 2009) and to improve students’ skills in critical thinking and analysis (Treyce et al., 2011). In addition, the inquiry-based laboratory was used to enhance students’ understanding and engagement in experimental design (Hofstein and Lunetta, 2004; Howard and Miskowski, 2005; Bugarcic et al., 2012); to increase student excitement about and motivation for engaging in research (Knudson et al., 2010); and to promote curiosity, creativity, and critical thinking (Zion and Sadeh 2007; Casotti et al., 2008; Moskovitz and Kellogg, 2011). The research-based laboratory was also shown to bring about greater learning investment in and excitement about the experiment when compared with non–research-based projects (Brame et al., 2008). It provided students with immense benefits over traditional laboratory experiences, and even over inquiry-based laboratory experiences (Weaver et al., 2008; Brownell et al., 2012). Many derivatives and hybrids have originated from the three approaches, such as guided-inquiry learning (Spiro and Knisely, 2008) and investigative and cooperative learning (Seifert et al., 2009), as well as integrated teaching and research (Kloser et al., 2011). These approaches offer varying levels and forms of autonomy support for students’ cognitive engagement and learning motivation. However, giving students autonomy to carry out experiments that interest them creates many challenges and a heavy workload for technical and support staff, particularly in large classes. It is generally agreed that increased benefits for students run parallel to increased difficulties for implementation from cookbook-style to project-based, inquiry-based, and research-based laboratories (Oliver-Hoyo et al. 2004; Roehrig and Luft, 2004; Weaver et al., 2008; Furtak and Kunter, 2012). The difficulties largely stem from logistics and little incentive for faculty to dedicate much time to teaching (Anderson et al., 2011; Kloser et al., 2011).

Therefore, continuing efforts must be dedicated to developing new pedagogic strategies that increase student autonomy but also remain feasible for educators constrained by large class sizes and modest budgets. In this study, we applied the concept of gene mutation from genetics to transform a cookbook-based molecular genetics lab exercise into a hypothesis-based inquiry science lab. In this new approach, students enjoy the freedom to “mutate” teacher-designed experiments to test their hypotheses and interpret experimental data in a written report. We prefer the word “mutate” to “alter” because the different types of mutations in genetics generate constructive ideas on how an experiment can be redesigned. In this approach, which we named “mutation-based learning” (MBL), we aim to enhance learner autonomy and scientific inquiry skills. The feedback from different batches of students over three semesters shows that this innovative approach has successfully improved student engagement and motivation. The MBL approach also provides a new venue for students to develop their scientific inquiry skills by studying case scenarios as scientists do. Furthermore, the implementation of MBL does not incur significant increase of labor and financial costs, thus making it practical.

**METHODS**

**Module and Laboratory Contents**

This study involved a first-year undergraduate module in molecular genetics, which is required for life sciences majors in the Faculty of Science at the National University of Singapore (NUS). The module is taught over 65 h and consists of lectures (36 h), tutorials (15 h), and laboratory sessions (14 h) within one semester. Class sizes varied between 200 and 280 students. The module contents include three parts. The first part covers DNA structure; replication; and gene transcription, translation, and regulation. The second part focuses on cell division; chromosome transmission and organization; and gene transfer, mapping, and recombination. The last part deals with Mendelian and population genetics at the molecular level. Each part, which is taught by a different lecturer, has a continuous assessment component contributing 15% to a student’s final module score.

The laboratory contents include genomic DNA extraction and agarose gel electrophoresis (practical one, abbreviated as P-one), plasmid DNA purification and transformation (P-two), and basic bacterial manipulation (P-three). P-one and P-three were carried out with a conventional cookbook approach, while P-two was conducted using the innovative MBL approach. Both approaches are explained in the following section. P-two was chosen for MBL because it started in the third week, while P-one and P-three commenced in the first week. This allowed students more time to know their classmates before forming groups and to familiarize themselves with the laboratory equipment and requirements for the MBL approach. The laboratory contents are widely connected with the second part of the lecture contents, such as genome organization and gene transfer and recombination. The assessment of laboratories contributes 10% to student’s final module score. This 10% is further divided into three parts: 2% for proposal writing, 2% for laboratory performance, and 6% for laboratory reports.

**Conventional Cookbook Approach versus Mutation-Based Learning Approach**

With the cookbook-style approach, a teacher-designed, step-by-step protocol (see standard protocol in the Supplemental Material) and related references were distributed to students at the beginning of each semester. Students were asked to read the materials before they attended the lab session. During practical sessions, students carried out experiments in pairs, strictly following the protocol and teaching assistants’ (TAs) guidance.
In the MBL approach, students were also given the standard protocol, but they were asked to look for references and to “mutate” the standard protocol with a written proposal to explain the rationale for their mutations. The method to mutate an experimental protocol is conceptually similar to the method used for gene mutations. Students could mutate the experiments by: deleting certain experimental steps in the standard protocol (namely deletion mutation), adding extra steps (insertion mutation), reversing the order of experimental steps (reverse mutation), or replacing certain steps (substitute mutation). TAs and instructors subsequently assessed students’ mutated protocols and proposals. Freedom for students to develop and implement their proposals was usually granted, unless there were safety concerns and logistical difficulties. During the scheduled lab sessions, students in groups of four performed the P-two in parallel, following both standard and mutated protocols. Additional time needed in the lab (e.g., for preparation of additional buffer or extra efforts for data collection) was available by appointment. After the lab session, the students could share, compare, analyze, and discuss their results in a group, but they were required to write individual reports.

All students experienced the two approaches, which enables comparison of each individual’s learning experiences via questionnaires. However, they carried out different experiments using different approaches. It is not educationally effective to ask students to repeatedly perform the same experiments via the two approaches. Logistically, it would be very difficult to have half of the class doing cookbook style, while the other half is using the MBL approach during a week of laboratory sessions. It would also be unfair to subject the students to a “controlled” cookbook approach as soon as any indication that the MBL approach is better arises. Thus, quantitative comparisons of student scores from the two lab approaches in the laboratory were not present.

**Student Autonomy Supports in the MBL Approach**

Various choices were offered in P-two to support student autonomy. The three categories of autonomy supports proposed for classroom teaching by Stefanou *et al.* (2004) were adopted but redefined in the context of laboratory activities (Table 1).

<table>
<thead>
<tr>
<th>Category of autonomy</th>
<th>Student’s self-determined activities</th>
</tr>
</thead>
</table>
| **Organizational autonomy** | • Form a group (relatedness support is needed here)  
• Decide when/where to discuss their project  
• Decide their roles in the project, such as uploading materials, communications with TAs, editing proposal, etc.  
• Decide when to prepare additional reagents and to collect extra data if needed |
| **Procedural autonomy** | • Develop their own experimental protocol  
• Make a time plan for additional experimental activities  
• Learn required experimental skills from TA (competence support also needed)  
• Perform experiments collaboratively as planned |
| **Cognitive autonomy** | • Look for relevant references and provide literature review  
• Identify a question of interest to the group  
• Make a hypothesis based on the experimental contents (case scenario)  
[Permission to proceed required at this point]  
• Collect data on a group decision  
• Individually process, present, and interpret data  
• Write an individual report |

### Assessment Rubrics

**Proposal of Mutation.** Students were asked to submit their proposals 1 wk before the first lab session for P-two. The proposal was limited to two pages and was evaluated by TAs. Students’ competence in formulating hypotheses and reasoning was the main consideration for assessment among many scoring points (Table 2). A maximum of 8 points was given to a group proposal submitted by four students. Depending on the level of individual contribution to the proposal, the students would then decide how to share out the total points assigned by the TA among the group members. When an individual contributed more to the proposal, he or she would be entitled to a larger portion of the total points. This exercise simulates the delicate collaborative and recognition behavior that research scientists engage in on a regular basis. The TA and/or instructor would help to make a final decision regarding the point distribution only when an agreement could not be reached among the group members.

**Implementation of Proposed Mutation.** Students’ performance during the lab session was monitored and evaluated. Each TA guided a group of 16–24 students, demonstrating technique skills and explaining key principles underlying the experiments, while instructors provided short overviews of important concepts and overall supervision for the entire class to encourage active participation and learning. TAs evaluated the performance of each student for each lab session based on the following rubrics, with major emphasis on student participation and conscientiousness (Table 3).

**Laboratory Report.** Students were instructed to write a laboratory report within four pages consisting of introduction, materials and methods, results, and discussion sections. Experimental results were shared among the group members, but each member had to write an individual report to ensure everyone was trained in scientific report writing. As they were first-year students, detailed instructions on writing a scientific report were given during a special tutorial. Their reports should present a clear comparison between the experiments conducted according to standard protocols (control) and mutated protocols (experimental group). Students had to submit soft copies of their reports to an online folder before the deadline. A 50% deduction was applied to any student who submitted their report after the deadline.
Table 2. Assessment rubrics for student proposals

<table>
<thead>
<tr>
<th>Proposal content</th>
<th>Student goals/desired outcomes</th>
<th>Score points</th>
<th>Lose points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction (40%)</td>
<td>1. Search and read relevant references and identify questions by themselves, showing enhanced cognitive autonomy</td>
<td>Briefly review the relationship of plasmid conformation and transformation efficiency; explain interest in a particular mutation and its significance and connection to the module/previous knowledge</td>
<td>Unpersuasive, disjointed, or unclear reasoning</td>
</tr>
<tr>
<td></td>
<td>2. Produce well-specified goals and objectives for the mutations, showing independent thought and ownership of learning</td>
<td>• Methods commonly used for DNA delivery, references cited</td>
<td>Fails to use references and/or previous knowledge</td>
</tr>
<tr>
<td></td>
<td>Example: A mutation is to substitute a circular plasmid with a linearized plasmid.</td>
<td>• Concept of transformation efficiency and the factors affecting the efficiency, references cited</td>
<td>Lacks or supplies an insufficient or wrong description of the following:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Why the substitution is of interest</td>
<td>• The concept of transformation efficiency</td>
</tr>
<tr>
<td>Hypothesis (10%)</td>
<td>1. Learn how to ask a scientific question and write a scientific hypothesis</td>
<td>Has scientific merit; is testable and provable</td>
<td>Poor hypothesis, as in the following:</td>
</tr>
<tr>
<td></td>
<td>2. Foster curiosity and creativity, enhance cognitive autonomy</td>
<td>• Well-phrased intellectual guess</td>
<td>• Linearized plasmid is cleaved by bacterial enzymes (without subsequent test to prove it)</td>
</tr>
<tr>
<td></td>
<td>The example continued:</td>
<td>A hypothesis being similar to the following:</td>
<td>• Linearized plasmid takes longer time to enter bacterial cells, leading to lower transformation efficiency (it is difficult to measure the time and the hypothesis is therefore not testable)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Linearized plasmid DNA leads to lower transformation efficiency compared with circular plasmid via heat shock method</td>
<td>Outcome unknown or immeasurable or not observable, unclear what to be tested, no proper control</td>
</tr>
<tr>
<td>Prediction (10%)</td>
<td>1. Learn how to set controls and minimize the variables</td>
<td>Expected outcomes if experiments are done</td>
<td>Incomplete list of materials needed or the list does not match the needs, incoherence of protocol changes</td>
</tr>
<tr>
<td></td>
<td>2. Appreciate that reductionisms (i.e., reduce the complexity of testing) play an extremely significant role in scientific inquiry</td>
<td>Preferably, only one variable is tested</td>
<td>• Without quality check after cleavage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• A prediction tells the dependent and independent variables, being similar to the following:</td>
<td>• A wrong enzyme chosen</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Different delivery methods used, such as heat shock method for circular plasmid, while electroporation for linearized plasmid</td>
</tr>
<tr>
<td></td>
<td>3. Enhance organization and procedural autonomy</td>
<td></td>
<td>• Different buffers used (more than one variable)</td>
</tr>
<tr>
<td></td>
<td>The example continued:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutated protocol and needed materials (20%)</td>
<td>1. Develop student competence and know-how/ experimental skills</td>
<td>Clearly list what is required and highlight the mutated steps; the experiments, hypothesis and protocols agree with each other</td>
<td>Data are not useful for validation of prediction, or required data are not collected</td>
</tr>
<tr>
<td></td>
<td>2. Learn how to plan their experiments timely and logistically, ensure experiments are feasible</td>
<td>• Enzyme (e.g., EcoRI) required for linearizing the plasmid</td>
<td>Without or insufficient</td>
</tr>
<tr>
<td></td>
<td>3. Enhance cognitive autonomy</td>
<td>• A miniscale DNA purification kit required to purify the DNA</td>
<td>• Data to show the plasmid was linearized</td>
</tr>
<tr>
<td></td>
<td>The example continued:</td>
<td>• Data to show the amount and purity of plasmid used</td>
<td>• Data to show the amount of plasmid used</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• A set of agarose gels to examine the cleavage, etc.</td>
<td>• Data to show the transformation efficiency</td>
</tr>
<tr>
<td>Planned data collection and analysis (20%)</td>
<td>1. Plan experiments in a coherent manner, so the data can be meaningful to verify the hypothesis</td>
<td>Determine data to be collected and analysis to be conducted</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Think critically to define the controlled variables</td>
<td>Any additional efforts to collect data</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Enhance cognitive autonomy</td>
<td>• Measure the plasmid purity and quantity</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The example continued:</td>
<td>• Electrophoresis to check the efficiency of linearization</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Count the number of colonies</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Compare transformation efficiency</td>
<td></td>
</tr>
</tbody>
</table>
report submitted before a late submission deadline (within 2 d past the set deadline), and reports were not accepted at all after the late submission deadline. TAs were instructed to base their assessment of the reports mainly on critical thinking and logical reasoning, so not much emphasis was placed on the experimental results per se. To avoid bias, a TA did not mark the reports from students who were under his or her supervision.

Report scores were classified under five categories: introduction (10%), materials and methods (5%), results (40%), discussion (40%), and other (5%). The percentage for each category differed slightly in each semester. A report score was the sum of the points that a student earned in the five categories. The average of the report score was calculated by dividing the total scores of all reports by the total number of student submissions. The percentage scored for each category was also calculated by dividing points earned by students by the maximum points of the respective category. Report marking rubrics with an example are shown in Table 4.

**Feedback Collection.** Anonymous Likert questionnaires were administered to all participants after the entire lab course using forms 1 (Table 5) and 2 (Table 6). Form 1 was used for the first trial during semester 2 in the academic year 2010–2011 (abbreviated as AY1011). Form 2, which had additional questions, was used to achieve direct measurement of students’ learning and engagement in AY1112 semesters 1 and 2. Students were instructed to make a comparison between the mutated (P-two) and the conventional experimental work (P-one and P-three) in the same course. In addition, students were also encouraged to write comments on the module website or to send feedback to TAs and instructors by email.

Both forms 1 and 2 adopt the 5-point Likert scale to assess different levels of efforts/satisfaction of the students, from the lowest to the highest: strongly disagree (1 point), disagree (2 points), undecided (3 points), agree (4 points), and strongly agree (5 points). The sum score for each question in the survey form 2 was calculated with the formula: sum score = \( \sum (nL \times L) \), where \( L \) is the Likert level (1–5) and \( nL \) is the number of respondents at the corresponding Likert level. The “difference of sum scores” was calculated for each question by subtracting the sum score for the cookbook-based approach from the sum score for the MBL approach. The average score of a question was calculated by dividing the sum score by the total number of respondents \( (n = 320) \). A paired two-tailed Student’s \( t \) test was used for statistical analysis of 320 individual scores from the two approaches. \( P < 0.001 \) was considered as the significance level after a Bonferroni correction via the formula \( 1 - (1 - \alpha)^{1/n} \), in which \( \alpha = 0.01 \) and \( n = 13 \) (total number of survey questions in Table 6).

In addition, one 5-point Likert questionnaire (form 3 in the Supplemental Material) was used to collect students’ evaluations of TAs. It provided a measurement of TAs’ capabilities in implementing the teaching strategy and in providing students with competence and relatedness support in addition to autonomy support.

**RESULTS**

**Students’ Proposal**

The students designed a number of mutations of P-two and conducted the mutated P-two in many different ways. In contrast, the entire class performed the same experiments in P-one and P-three, following the standard protocols. Representative mutations that appeared in a few groups’ proposals were mostly due to dramatic changes of the standard protocol, resulting in too many variables and untestable conditions. In the P-two scenario, plasmid extraction and

<table>
<thead>
<tr>
<th>Performance</th>
<th>Goals/desired outcomes</th>
<th>Score points</th>
<th>Lose points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Safety (20%)</td>
<td>Be familiar with safety rules, knowing that safety issues are paramount and cannot be compromised under any circumstances</td>
<td>Comply with safety rules and wear proper protection attire</td>
<td>Wear improper attire such as slippers and shorts</td>
</tr>
<tr>
<td>On-schedule (20%)</td>
<td>Enhance organization autonomy, self-discipline, and teamwork attitudes</td>
<td>Arrive on time and manage to finish planned experiments efficiently</td>
<td>Late</td>
</tr>
<tr>
<td>Conduct (60%)</td>
<td>Enhance procedural autonomy and competence in skills</td>
<td>Know what to do</td>
<td>No clear working plan</td>
</tr>
<tr>
<td></td>
<td>Develop interpersonal skills, such as being collaborative, peer teaching, negotiation, and collaboration</td>
<td>Engaged and collaborative</td>
<td>Unable to complete experiments within time frame</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Appropriate use of equipment and apparatus</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conscientious</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Take notes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Play mobile devices</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chit-chat</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rash, careless, mishandle</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>materials</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No participation and data collection</td>
<td></td>
</tr>
</tbody>
</table>

*Table 3. Assessment rubrics of student performance in the laboratory*
Table 4. Scoring rubrics of laboratory reports

<table>
<thead>
<tr>
<th>Content</th>
<th>Goals/desired outcomes</th>
<th>Score points</th>
<th>Lose points</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Introduction (10%)</strong></td>
<td>1. Promote extensive reading of relevant references&lt;br&gt;2. Generate curiosity about “what I want to know”&lt;br&gt;3. Correlate known knowledge to unknown area&lt;br&gt;The example in Table 2 continued:</td>
<td>Be consistent with the Introduction of the proposal&lt;br&gt;Provide the most salient background closely related to the proposed mutation&lt;br&gt;Comprehensive review of mechanisms of heat shock transformation&lt;br&gt;More information about factors affecting transformation efficiency, especially as related to DNA conformation</td>
<td>Inconsistent with the proposal&lt;br&gt;No references cited&lt;br&gt;Irrelevant materials used, such as&lt;br&gt;Horizontal gene transfer&lt;br&gt;Relationship of transformation and spread of antibiotic-resistant bacterial strains&lt;br&gt;Bacterial pathogenicity/virulence</td>
</tr>
<tr>
<td><strong>Materials and methods (5%)</strong></td>
<td>1. Emphasis on communicative purpose, being aware that there is no need to repeat something well known among audience&lt;br&gt;2. Be sufficient to allow the experiments to be repeated&lt;br&gt;3. Be technically competent&lt;br&gt;The example continued:</td>
<td>Be brief and consistent with the proposed alteration of standard protocol&lt;br&gt;Highlight “the mutated part”&lt;br&gt;Avoid redundancy with the content in the lab manual [This differs from published scientific reports.]&lt;br&gt;As simple as one sentence for plasmid extraction: “Plasmid was extracted using a High-Speed Plasmid Mini Kit (Geneaid Biotech) according to the manufacturer’s manual.”&lt;br&gt;An equal amount of linearized and circular plasmid was used to transform E. coli cells by a heat shock method</td>
<td>Copy and paste the details from lab manual and proposal&lt;br&gt;A long and wordy comparison of standard and mutated protocols&lt;br&gt;List experimental steps like a cookbook-style manual&lt;br&gt;List buffers, reagents, and required volumes already shown in the laboratory manual&lt;br&gt;Wrong information about the plasmid extraction kit, competent cells, and plasmid used</td>
</tr>
<tr>
<td><strong>Results (40%)</strong></td>
<td>1. Learn how to process and present data scientifically&lt;br&gt;2. High standard of scientific integrity, being aware of data misconduct, such as fabrication, manipulation, and falsification&lt;br&gt;3. Maintain originality and avoid plagiarism&lt;br&gt;4. Comparing and contrasting especially crucial in MBL&lt;br&gt;The example continued:</td>
<td>High clarity and in logical order&lt;br&gt;Contents precise and consistent in figures, tables, legend, and text&lt;br&gt;Plasmid concentrations and qualities obtained from different experimental designs&lt;br&gt;Gel photos showing linearized plasmid&lt;br&gt;Number of colonies obtained from different designs&lt;br&gt;Transformation efficiency under different conditions&lt;br&gt;Analysis to show differences of results between different designs</td>
<td>Fabricating or selecting data with bias&lt;br&gt;Presenting raw data without organization and processing&lt;br&gt;Inconsistent or disorganized&lt;br&gt;Redundant&lt;br&gt;Supply no or only partial results&lt;br&gt;Cover or ignore unexpected results&lt;br&gt;Data presented in a messy way&lt;br&gt;Lack or give insufficient description of table or figure&lt;br&gt;Without analysis, only number is presented&lt;br&gt;Redundant figures and tables</td>
</tr>
<tr>
<td><strong>Discussion (40%)</strong></td>
<td>1. Compare, evaluate, and interpret the results critically&lt;br&gt;2. Logic is clear and the statement convincing</td>
<td>Critically and scientifically interpret outcomes from mutated and standard protocols&lt;br&gt;Comprehensive explanation of unexpected results&lt;br&gt;Conclusion drawn based on data</td>
<td>Discussion does not lead to a conclusion or correlate to the proposal and the results&lt;br&gt;Simply repeat introduction, results, a principle, or a theory without personal insights given</td>
</tr>
</tbody>
</table>
transformation, we did not expect students to make significant findings. However, the references, questions, hypotheses, and mutated protocols in students’ proposals demonstrated that their autonomy and inquiry skills were enhanced.

In addition, we seldom encountered difficulties when supporting proposed mutated experiments, because the substituted mutations only needed additional materials that usually consisted of very common chemicals. Logistical support required for other types of mutations are the same as those in a cookbook-style approach.

**Lab Performance**

Based on the TAs’ observation, students were much more serious when doing P-two than they were when doing P-one and P-three. Students were very conscientious and spent a longer time completing experiments in MBL than they had anticipated. This was especially common when they prepared the reagents. They were not practically proficient in calculating molar concentration, adjusting pH, and weighing very small amounts (e.g., in milligrams or micrograms) of chemicals, although they were often amazed at advanced infrastructures and instruments. How plasmid DNA could be transferred into bacterial cells by such a short heat shock process often triggered contemplation. They collaborated as a team and engaged in experimenting. They were impressed by their observation of DNA bands on the UV trans-illuminator, gel electrophoresis, and measurement of plasmid DNA concentration and quality by NanoDrop. Different colors of colonies from *Escherichia coli* cells with and without plasmid (pUC18) made the relationship between genotype and phenotype distinct for them, thus reinforcing the relatedness with lecture content. The students also observed safety rules and were able to keep paper and electronic records of experimental steps, observations, and data. Only a small cadre of students (∼1%) showed any reluctance in spending additional time on the mutated experiments.

### Table 4. Continued

<table>
<thead>
<tr>
<th>Content</th>
<th>Goals/desired outcomes</th>
<th>Score points</th>
<th>Lose points</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Synthesize or create new information or ideas for further study</td>
<td>Discuss how the ratio of plasmid amount to the number of competent cells affects the transformation efficiency</td>
<td>Overreached conclusion, such as:</td>
<td></td>
</tr>
<tr>
<td>The example continued:</td>
<td>Interpret the different transformation efficiencies resulting from circular and linear plasmid</td>
<td>Irrelevant: e.g., importance of “mutated” experiment to study genetics, how the plasmid contributed to the antibiotic’s resistance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Predict what are the possible factors leading to the differences; references must be cited in support</td>
<td>Goes beyond published and/or experimental data, too much speculation, e.g., plasmid interacts with some unknown proteins, affecting the transformation efficiency</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Explain factors causing the errors or deviations from the results expected, such as efficiency of enzyme cleavage, plasmid purity, etc.</td>
<td>Reiterate principles without correlation to the results</td>
<td></td>
</tr>
<tr>
<td>Others (5%)</td>
<td>1. A paper must be presentable and readable</td>
<td>References in a unified format</td>
<td>References not well organized</td>
</tr>
<tr>
<td></td>
<td>2. Encourage creativity</td>
<td>Presentable figures and tables</td>
<td>Figures and tables wrong size, alignment, and/or position</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Paper is easy to read</td>
<td>Poor grammar and spelling errors</td>
</tr>
<tr>
<td></td>
<td>Bonus (maximum of 5 points) for creativity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5. Survey form 1 used for students to compare the mutation-based approach with the conventional cookbook approach in semester 2 of AY1011**

<table>
<thead>
<tr>
<th>Survey questions</th>
<th>Student numbera</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>% of (5 + 4)</th>
<th>% of (2 + 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: The mutation approach is more challenging and stimulating.</td>
<td>57 147 52 5 2</td>
<td>77.6</td>
<td>2.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q2: The mutation approach enhances a sense of ownership for my learning.</td>
<td>59 143 55 4 2</td>
<td>76.8</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q3: The mutation approach gives me freedom to test my own idea/hypothesis.</td>
<td>74 149 34 4 2</td>
<td>84.8</td>
<td>2.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q4: The mutation approach enhances my interest in molecular genetics lab.</td>
<td>44 147 62 7 3</td>
<td>72.6</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q5: The mutation approach improves my critical-thinking skills.</td>
<td>51 160 47 4 1</td>
<td>80.2</td>
<td>1.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q6: The mutation approach improves my ability to communicate and work with teammates.</td>
<td>55 164 42 2 0</td>
<td>83.3</td>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q7: Overall, mutation-based learning is more effective.</td>
<td>55 156 48 4 0</td>
<td>80.2</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aThe scores from 5 to 1 represent different agreement levels: strongly agree, agree, neutral, disagree, and strongly disagree, respectively. At the right side of each question, it shows the student numbers at different levels of agreement. The % of (5 + 4) represents the percentage of students in “agree and strongly agree”; the % of (2 + 1) represents the percentage of students in “disagree and strongly disagree.”
In addition, the students were impressed with TAs competences in technical skills and knowledge, giving TAs an average score of 4.4 out of 5 (data 3 in the Supplemental Material), indicating satisfaction with the competence support from TAs. Overall, the MBL laboratory not only provides autonomy support, but also competence and relatedness support. One quote from a student follows:

"The "mutations" introduced by virtue of modifying experimental protocol were very innovative and interesting. Though there were constraints on the kind of "mutations" we could implement, we were really required to think hard about the consequences of each action, rather than just mindlessly follow a protocol and answer some questions thereafter. I believe this kind of thinking is beneficial to our development as scientists or even science students, as it is necessary for us to practice how to make hypotheses and carry out experiments to determine the final result."

### Lab Reports

Reports were first scrutinized through the Turnitin program for plagiarism detection. Before we conducted MBL, we had identified a few cases (<1%) in the previous years in which students copied a classmate's or a senior's report. Further investigation revealed that a plagiarism offender often made an excuse that he or she referred to but did not copy another report of a student who had done the exactly same experiments. After the implementation of MBL, no plagiarism was identified. The average score of the lab reports was 82 out of 100 in the three semesters. As for the average scoring percentages of the five categories in the report marking rubrics, they were 85, 90, 88, 81, and 90% for the introduction, materials and methods, results, discussions and others, respectively.

### Feedback from Students

A survey using form 1 was administered after the first trial of MBL. All feedback was manually inspected. The number of students at different levels of agreement with the seven survey questions is supplied in Table 5. A substantial proportion of students (72.6–84.8%) were of the same mind (strongly agreed or agreed) that MBL was better than the cookbook-style learning for all seven questions in form 1. In contrast, only 0.8–3.8% of students thought otherwise (disagreed and strongly disagreed). The highest satisfaction was given to question 3 (Q3), regarding the freedom of idea testing, while the lowest went to Q4, which was about interest enhancement.

Encouraged by the promising results from AY1011 (Table 5), MBL was deployed in the following two semesters in AY1112. A total of 188 and 171 survey forms were gathered after the laboratory course in semesters 1 and 2, respectively. After a manual check to remove partially filled and illegible forms, 156 and 164 forms, representing 77 and 61% of student cohorts in semesters 1 and 2, respectively, were used for data collection and analysis. The total number of students who strongly agreed and agreed with all 13 questions were larger when students carried out P-two via MBL than the number of students who did the experiments via the conventional approach in the two semesters (P-one and P-three; Table 6). The greater difference between the total scores of MBL and conventional learning was associated with Q2, Q5, Q8, and Q9, with different scores of 124, 101, 101, and 195, respectively. The smaller difference appeared to be related to Q1 and Q3 (Table 6). The differences between the learning activities in the two approaches are summarized in Table 7.

The different levels of agreement from “strongly disagree” to “strongly agree” were assigned a score from one to five. The average feedback score for each question was calculated for the cookbook-based approach from the score for the MBL approach.

<table>
<thead>
<tr>
<th>Survey questions</th>
<th>MBL approacha</th>
<th>Cookbook-style approachb</th>
<th>Difference of sum scoresb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: I make efforts to understand the experimental design before the lab.</td>
<td>75</td>
<td>51</td>
<td>24</td>
</tr>
<tr>
<td>Q2: I look for additional materials related to the experimental contents.</td>
<td>166</td>
<td>114</td>
<td>52</td>
</tr>
<tr>
<td>Q3: I am eager to try the experiments.</td>
<td>58</td>
<td>61</td>
<td>51</td>
</tr>
<tr>
<td>Q4: I am curious to know what will happen to the experiments.</td>
<td>7</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Q5: I discuss the experiments before the lab class.</td>
<td>51</td>
<td>51</td>
<td>0</td>
</tr>
<tr>
<td>Q6: I discuss the experiments after the lab class.</td>
<td>71</td>
<td>48</td>
<td>23</td>
</tr>
<tr>
<td>Q7: The exercise in the molecular genetics lab is challenging and stimulating.</td>
<td>73</td>
<td>64</td>
<td>9</td>
</tr>
<tr>
<td>Q8: It enhances a sense of ownership/responsibility for my learning.</td>
<td>86</td>
<td>48</td>
<td>38</td>
</tr>
<tr>
<td>Q9: It gives freedom to test my own idea/hypothesis.</td>
<td>110</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>Q10: It enhances my interests at molecular genetics/biology lab.</td>
<td>79</td>
<td>55</td>
<td>24</td>
</tr>
<tr>
<td>Q11: It improves my critical-thinking skills.</td>
<td>80</td>
<td>52</td>
<td>28</td>
</tr>
<tr>
<td>Q12: It improves my ability to communicate and work with teammates.</td>
<td>93</td>
<td>68</td>
<td>25</td>
</tr>
<tr>
<td>Q13: Overall, learning in the lab is effective.</td>
<td>90</td>
<td>66</td>
<td>24</td>
</tr>
</tbody>
</table>

aThe number of students at different levels of agreement to the questions; the scores from 5 to 1 represent strongly agree, agree, neutral, disagree, and strongly disagree, respectively.
bSum score = \( \sum (nL \times L) \), where \( L \) is the Likert level (1–5) and \( nL \) is the number of respondents at the corresponding Likert level. The difference of sum scores was calculated for each question by subtracting the sum score for the cookbook-based approach from the score for the MBL approach.
Table 7. Learning activities in conventional cookbook-style and MBL approaches

<table>
<thead>
<tr>
<th></th>
<th>Conventional cookbook-style learning approach</th>
<th>MBL approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prelab</td>
<td>Spend little effort to understand lab contents</td>
<td>Make effort to redesign/mutate experiments, including reference review, group discussion, proposal writing, and preparation of chemicals and buffers</td>
</tr>
<tr>
<td>In lab</td>
<td>Follow standard instructions, conduct experiments in an orderly manner without deviations, get anticipated results for most experiments, and learn technique skills and underlying principles</td>
<td>Conduct experiments facilitated by teaching assistants, experience uncertainties and a sense of ownership, get various and sometimes unanticipated results, learn additional technique skills/knowledge by doing</td>
</tr>
<tr>
<td>Postlab</td>
<td>Individually process data and write a report, reiterate underlying principles in their own words</td>
<td>Share and discuss results with group members, compare and analyze results from different approaches, cope with unanticipated data, re-examine experimental design, write a report (as an individual, not as part of a group), form a conclusion to support or falsify a hypothesis</td>
</tr>
</tbody>
</table>

for both approaches were given to Q4, regarding curiosity (4.23 in MBL vs. 4.02 in the conventional approach), while the lowest scores (3.45 vs. 3.14) were given to Q5, regarding the efforts spent to discuss experiments before the lab class. The scores for the overall effectiveness of laboratory learning (Q13) were 4.15 versus 3.96. Statistical analysis showed the average scores for 11 survey questions are significantly higher ($P < 0.001$) for the MBL approach than for the conventional approach. The feedback scores from only two survey questions (Q1 and Q7) did not lead to such significance ($P > 0.001$).

**DISCUSSION**

A gene mutation is a change in the DNA sequence that may be beneficial, detrimental, or have no effect on an organism. Well-designed gene mutation has been an important approach for studying gene function in biology. We expected that a well-designed “mutation of experimental design” could be used to explore scientific questions under-
scores improved significantly through MBL, as evidenced by significantly higher scores of Q2, Q5, Q8, and Q9 (Table 6 and Figure 1). The active and engaged behaviors, such as reading additional references and discussing experiments with peers and TAs, manifest intrinsic motivational development, which is usually correlated to higher learning achievements (Elliot, 1999; Russell and French, 2002; Niemiec and Ryan, 2009). To foster stronger self-motivated and independent learning further, we may need to examine other factors, such as creating more attractive experimental scenarios (Katz and Assor, 2007) and adopting new grading methods for the module (Shim and Ryan, 2005).

Systematic analysis of students' reports revealed that students were able to make critical comparison of experimental data from both teacher-designed and mutated protocols. Their reports adopted the same format for papers in scientific journals. They generally wrote well, with an average score of 82 points out of 100. The scoring percentages for introduction (85%), materials and methods (90%), results (88%), and discussion (81%) show that students faced more challenges in writing the introduction and discussion sections. The difficulty in writing these two sections is a common challenge for scientists. The overall high scores may indicate that students are capable of writing reports in a scientific journal format if they have genuine experimental scenarios and data.

Importantly, the greatest difference of total scores was generated from Q9 (Table 6), which indicates that students were highly satisfied with the freedom to test their own hypotheses in the MBL. This freedom is critical for enhancing autonomy and ownership of learning. A great number of diverse mutations and hypotheses demonstrate various interests, understanding, and creative-thinking skills among students (data 2 in the Supplemental Material). This diversity and the broad choices are different than project-, inquiry-, and research-based laboratory exercises, in which available projects are usually limited and designed for a small class size (Weaver et al., 2008).

The students' proposals were very diverse. Some proposals in MBL were creative, innovative, and scientifically sound. Others were driven by curiosity, because students doubted what is stated in the textbook. Some students ventured to make new findings, while others were very conservative in securing a set of data for reporting. Not surprisingly, some students made illogical hypotheses. In high school or conventional laboratories, students are usually misled by “foolproof” experiments that make science appear to be straightforward and always yielding anticipated results. In contrast, MBL allowed students to know the complexity of science research and nature of science by doing. Students were nudged or motivated to look for references, discuss with group members, and interpret both anticipated and unanticipated results. All these processes eventually provide the requisite training in scientific inquiry skills.

All students knew that the standard protocol had been optimized and worked well for experimental purposes. Some students turned their focus to understanding the science behind the each experimental step. Their mutations were proposed to verify the importance of the particular step or the function of chemicals. Nevertheless, a comparison of results from different approaches is still helpful to improve their critical-thinking skills and understanding of experimental principles (Spiro and Knisely, 2008). Overall, only a small percentage of proposals were put forward to improve the experimental design, and very few led to potential improvement of the standard protocols, such as using plasmid mass instead of volume to optimize transformation efficiency.

In addition to promoting student autonomy, the MBL approach provides opportunities for teaching scientific integrity. During student consultation, a substantial number of students did not get the data they expected. Most of these unanticipated data were derived from students venturing in innovation or improvement of the standard protocol. They felt confused and did not know how to deal with the unanticipated or conflicting data. Some were liable to take a shortcut, simply choosing the data supporting their hypotheses and forgoing those unanticipated data, so finishing their reports became easy. Some even intended to modify the data to make them appear better. To ensure scientific integrity, we repeatedly reminded students to record all data without bias during all lab sessions, not to group their data into “bad” and “good” categories, and not to copy data from others. They were encouraged to organize and interpret the data critically and to prepare their reports independently. We highlight the value of learning from both successes and failures, and reward such learning in assessment by focusing on critical analysis and logical interpretation of results rather than on data alone. This may promote a better attitude toward science. Because the students made various mutations, the final report from each student was unique. This is distinctly different from the outcomes of the conventional laboratory, in which all students conduct the exact same experiments. In other words, MBL prevents students from a “copy-and-paste” approach at the onset of their experiments; instead, unique proposals are raised by autonomous learners.

In summary, while the conventional cookbook style and recently reformed approaches work in many ways to achieve certain pedagogic goals, the MBL approach supports student autonomy to acquire scientific inquiry skills and fosters intrinsic motivation and a sense of ownership of learning. It engages students in many of the same activities and thinking processes used by scientists. It is similar to an authentic inquiry process in which questions are not defined, experiments are not predesigned, and data are not provided to students (Buck et al., 2008). However, MBL has a few distinguishing features. First, it provides a reference (teacher-designed experiment) to explicitly guide students in designing experiments of interest to them. Second, the concept of gene mutation also encourages constructive ideas on how students can redesign the experiment. Third, students collect real data from comparable experiments to practice scientific writing. Fourth, logistical cost is limited. Finally, MBL can be implemented in large experimental classes. While some features may pare down complete autonomy, MBL offers competence support that is beneficial for first-year undergraduate students in their acquisition of scientific inquiry skills.

ACKNOWLEDGMENTS
I thank Janet Chan-Wong Swee Moi for her critical reading of the manuscript. I also thank all the lecturers, teaching assistants, and students for their active support and participation. This research is supported by Teaching Enhancement Grant (C-154-000-039-001) from the National University of Singapore (NUS).
REFERENCES


Bioinformatics Education in High School: Implications for Promoting Science, Technology, Engineering, and Mathematics Careers

Dina N. Kovarik,*, Davis G. Patterson,† Carolyn Cohen,‡ Elizabeth A. Sanders,† Karen A. Peterson,§ Sandra G. Porter,∥ and Jeanne Ting Chowning*†

*Northwest Association for Biomedical Research, Seattle, WA 98119; †University of Washington, Seattle, WA 98195; ‡Cohen Research & Evaluation, Seattle, WA 98115; §National Girls Collaborative Project & EdLab Group, Lynnwood, WA 98036; ∥Digital World Biology, Seattle, WA 98107

Submitted November 10, 2012; Revised March 7, 2013; Accepted April 4, 2013

We investigated the effects of our Bio-ITEST teacher professional development model and bioinformatics curricula on cognitive traits (awareness, engagement, self-efficacy, and relevance) in high school teachers and students that are known to accompany a developing interest in science, technology, engineering, and mathematics (STEM) careers. The program included best practices in adult education and diverse resources to empower teachers to integrate STEM career information into their classrooms. The introductory unit, Using Bioinformatics: Genetic Testing, uses bioinformatics to teach basic concepts in genetics and molecular biology, and the advanced unit, Using Bioinformatics: Genetic Research, utilizes bioinformatics to study evolution and support student research with DNA barcoding. Pre–post surveys demonstrated significant growth (n = 24) among teachers in their preparation to teach the curricula and infuse career awareness into their classes, and these gains were sustained through the end of the academic year. Introductory unit students (n = 289) showed significant gains in awareness, relevance, and self-efficacy. While these students did not show significant gains in engagement, advanced unit students (n = 41) showed gains in all four cognitive areas. Lessons learned during Bio-ITEST are explored in the context of recommendations for other programs that wish to increase student interest in STEM careers.

INTRODUCTION

Careers in science, technology, engineering, and math (STEM) command higher salaries, make significant contributions to the economy, and provide paths out of poverty, yet student interest in STEM careers is on the decline (Lowell et al., 2009; University of the Sciences [USciences], 2012). When high school students are asked to explain their lack of interest, 18% cite a lack of knowledge about health science careers, while others either blame poor preparation in high school (16%) or an inability to do well in these subjects (21%; USciences, 2012). A recent review of the literature concluded that “high school appears to be a key point at which young people’s impressions of science influence their future career decisions” (Subotnik et al., 2010). Understanding how high school students become aware of STEM career options, how educators can help students translate awareness into pursuit of STEM careers, and how to provide students with the support and skills they need to succeed are crucial elements for ensuring the future of our STEM workforce. According to the U.S. Department of Labor’s Bureau of Labor Statistics, life sciences and other STEM professions are expected to grow at a faster rate than non-STEM professions (U.S. Department of Labor, 2007), spurred on by both the fast pace of technological innovations and the changing
demographics of the U.S. population (Stine and Matthews, 2009). Consequently, national attention is focusing on the need to better promote STEM careers among young people (National Research Council [NRC], 2002; President’s Council of Advisors on Science and Technology [PCAST], 2010). We present here an investigation into how the Bio-ITEST program offers a teacher professional development model and student classroom experiences that seek to increase teacher and student awareness of and interest in bioinformatics-related STEM careers.

Cognitive-Behavioral Building Blocks of Career Development

A large body of work addresses various cognitive-behavioral processes that lead to career choice (or discouragement), including the concepts of awareness, self-efficacy (and closely related notions such as autonomy, confidence, proficiency, and competence), engagement (or involvement), and relevance (Bandura, 1994; Blustein and Flum, 1999; Dorsen et al., 2006). Though different conceptual frameworks use varying terminology, the approaches are complementary. These cognitive processes, occurring internally and mediated by both intrinsic and extrinsic motivators, result in external behaviors whose consequences in turn further reinforce or detract from emerging career choices.

Awareness of STEM careers is an essential precondition for engagement, self-efficacy, and a sense of relevance to develop; however, students often have limited understanding of available careers and requirements for success (Dorsen et al., 2006). Student awareness includes knowledge and appreciation of the required skills, education, and work/life issues associated with a variety of STEM careers. Tai et al. (2006) demonstrated that early expectations of a career in science are a potent predictor, independent of academic preparation, of later STEM career choice, emphasizing the need for early exposure to and encouragement in the pursuit of STEM careers.

Self-efficacy is defined by Bandura (1994) as “people’s beliefs about their capabilities to produce effects,” that is, to achieve particular results. Bandura further posits that, as people prepare for careers, perceived self-efficacy is the foundation of cognitive, self-management, and interpersonal skills informing career choice and success. A large body of work supports Bandura’s theory as it relates to achievement in STEM fields (Hackett and Betz, 1989; Lent et al., 1991; Mau, 2003; Zeldin et al., 2008). Career exploration activities can raise awareness and at the same time foster a sense of self-efficacy and ownership that becomes intrinsically motivating (Blustein and Flum, 1999).

Engagement with subject matter and STEM careers can be demonstrated by students showing interest in learning and experiencing more in the science classroom, including active participation in discussions and asking questions that go beyond the content presented. Engaging students in real-world research projects is a proven strategy to encourage interest in science careers (O’Neill and Calabrese Barton, 2005). Schneider et al. (1995) have also found that students who report high motivation and challenge in their schoolwork are more likely to engage in future educational opportunities.

Relevance is a concept that describes when students find a meaningful connection to STEM content or related careers and is a critical component in fostering the positive feelings associated with intrinsic motivation (Shernoff et al., 2003). Situations that require students to solve real problems serve to increase perceptions of relevance. Science relevance is typically measured in terms of student beliefs that science may be useful in everyday life and in the future (Siegel and Ranney, 2003).

While these conceptual elements of the cognitive-behavioral processes that lead to career choice are causally related to one another, each component may operate in multidirectional, simultaneous, or mutually reinforcing ways. For example, engagement can lead to increases in self-efficacy, heightening both awareness and a sense of relevance, while a sense of self-efficacy fostered by success can encourage further exploration of the subject matter, underscoring the relevance of the material to the student’s life and promoting greater engagement.

Bioinformatics and the Data-Driven Nature of Today’s Biology

Advances in data-intensive sampling methods, such as high-throughput DNA sequencing, proteomics, metabolomic characterization of complex biological samples, and high-resolution imaging of various living systems, have led to exponential growth in the amount of biological data available and rapid changes in how biological information is used. Bioinformatics can be defined as the application of computer science to biology and biomedicine. It is an interdisciplinary field that combines information technology, software engineering, and biology to analyze the massive data sets generated in biology today. Bioinformatics databases and analysis tools have become ubiquitous in modern biology: from DNA and protein comparisons to working with molecular structures and metabolic pathways, bioinformatics is integral to our understanding of biology. The need for individuals who can understand and analyze this wealth of information and utilize bioinformatics tools for data analysis has grown rapidly, with serious implications for our future STEM workforce. The NRC, in Building a Workforce for the Information Economy, notes that “the wealth of biotechnology-related data continues to expand, along with the need to analyze and understand it, and specialists in bioinformatics . . . are in great demand relative to supply . . . There is no sign that the demand for bioinformatics specialists is abating. Indeed, the demand will continue to grow rapidly, given estimates that as many as 40% of the biotechnology companies that survive will be selling information rather than products” (NRC, 2001, p. 328).

Despite the ubiquity of bioinformatics in biology today, these tools and concepts receive little attention in most high school science classes. This paper describes our efforts to address this need through Bio-ITEST: New Frontiers in Bioinformatics and Computational Biology. The Bio-ITEST program and curricula are designed to provide secondary science teachers with the knowledge, skills, and resource materials to engage their students in the newly developing fields of bioinformatics and related careers at the intersection of biology and information technology, encouraging student participation in these important new workforce areas.

Many bioinformatics tools used by scientists are freely available and can be readily implemented in high school settings with little to no up-front costs if student computers
are available. This provides students with the opportunity to use authentic bioinformatics and tools and databases utilized by practicing scientists. We predicted that the experience of using the same tools that scientists use in conjunction with sequential skill building from lesson to lesson would increase students’ sense of self-efficacy. Each Bio-ITEST lesson encourages career awareness by featuring a different STEM professional who uses bioinformatics in his or her work or whose work is made possible by bioinformatics. Bio-ITEST curriculum-development teachers selected genetic testing and evolution as unit topics, which they have found to be particularly engaging for high school students. These topics also provide narrative frameworks that can be investigated with bioinformatics tools and databases. We predicted that these topics would increase student perceptions of the relevance of unit content. In addition, both topics involve socio-scientific issues that can be explored through ethical analysis and discussion strategies. In our prior work, we have found that socio-scientific issues promote student interest in science content and foster critical-thinking skills (Chowning, 2005, 2009a, 2009b; Chowning et al., 2012). A feedback loop in which students document their increasing proficiency in using bioinformatics tools through résumé writing can serve to further increase self-efficacy and relevance. The concepts of awareness, self-efficacy, engagement, and relevance were also used to evaluate the effectiveness of the Bio-ITEST program by measuring changes in these attitudes relative to bioinformatics, bioethics, and related STEM careers.

Through bioinformatics curriculum development and teacher professional development, the long-term goals of the Bio-ITEST program are to increase teacher and student understanding of the application of information technologies to the biological sciences; the ethical implications of the acquisition and use of biological information; and the career possibilities in the fields of bioinformatics, computational biology, and related STEM careers.

**MATERIALS AND METHODS**

**Bioinformatics Curriculum Development Approach**

The Bio-ITEST curriculum development was led by the Northwest Association for Biomedical Research (NWABR), a 501(c)(3) nonprofit organization that has been fostering students and teachers in bringing the discussion of ethical issues in science into the classroom since 2000 (Miller, 2008; Chowning, 2005, 2009a, 2009b; Chowning et al., 2012). NWABR’s mission is to promote an understanding of biomedical research and its ethical conduct through dialogue and education. As part of this mission, NWABR has a demonstrated history of success in developing curricular materials and providing teacher professional development focused on the science and ethics of diverse topics, including HIV vaccines, stem cells, animals in research, and the nature of science (freely available at www.nwabr.org). NWABR connects the scientific and education communities across the Northwest and helps the public understand the vital role of research in promoting better health outcomes. NWABR’s curriculum-development process is informed by the principles of *Understanding by Design* by Wiggins and McTighe (1998), and “constructivist” perspectives that recognize that learners build their understanding based on prior experience and construct conceptual scaffolds upon which to integrate new learning. Teacher partners work with NWABR in all aspects of the curriculum development process, selecting central ideas, conceptualizing lessons, field testing, and sharing their experiences in teaching students and their knowledge of effective implementation strategies in the areas of state and national education standards.

Six experienced teachers from Washington and Oregon were recruited to provide the broad outlines of two bioinformatics-related curriculum units during the 2-wk 2009 Bio-ITEST curriculum development workshop. During the first week, teachers were immersed in bioinformatics and molecular biology through the use of wet-lab and computer activities, meetings with scientists, and tours of research facilities in the Seattle area. Computer activities included exploration of the molecular visualization program Cn3D (Wang et al., 2000) utilizing previously developed molecular structure activities that had been shown to significantly increase student learning among high school and college students (Porter et al., 2007). On the basis of these experiences, teachers chose overarching themes for the two bioinformatics units. The introductory curriculum focuses on genetic testing, which teachers believed would be relevant and engaging for students, particularly with new companies like 23andMe (Mountain View, CA) offering personalized, direct-to-consumer (DTC) genetic testing. This topic also offers the opportunity to address topics in molecular biology, bioinformatics, and ethics-related concepts. The advanced curriculum focuses on genetic research and evolution, utilizing DNA sequence data and bioinformatics tools to explore species relatedness.

Lesson materials were developed and further refined by Bio-ITEST staff following review by our advisory board of scientists and educators, and were then shared with teachers at semiannual professional development workshops (described below in Bioinformatics Teacher Professional Development). Workshop teachers then piloted and field-tested the lessons, providing written and oral feedback that informed additional lesson refinements in an ongoing and iterative process over a 2-yr period. Detailed site observations and teacher interviews conducted by the external evaluation team further informed revisions, particularly to the career materials contained in each unit. To ensure the accuracy and authenticity of Bio-ITEST curriculum materials, as well as to obtain feedback about lesson composition and flow, NWABR recruited additional bioinformatics experts to review both curriculum units. High school and college students were recruited to provide feedback on the content, flow, and usability of each lesson. All lessons were mapped to the Washington State K–12 Science Learning Standards (Office of the Superintendent of Public Instruction, 2010), the National Science Education Standards (NRC, 1996), and the A Framework for K–12 Science Education (NRC, 2011). All lessons are freely available as PDF documents with accompanying PowerPoint slides, sequence and structure files, and supporting animations. These materials can be downloaded from NWABR’s introductory bioinformatics curriculum Web page (www.nwabr.org/curriculum/introductory-bioinformatics-genetic-testing) and advanced curriculum Web page (www.nwabr.org/curriculum/advanced-bioinformatics-genetic-research).
Introductory Curriculum: Using Bioinformatics: Genetic Testing

The introductory bioinformatics curriculum, Using Bioinformatics: Genetic Testing, introduces students to a collection of bioinformatics tools and explores the ethical issues surrounding genetic testing. Students investigate the genetic and molecular consequences of a mutation in the Breast Cancer susceptibility 1 (BRCA1) gene, using the Basic Local Alignment Search Tool (BLAST) bioinformatics tool to compare DNA and protein sequences from patients with those of the BRCA1 reference sequence. Students then use Cn3D to visualize molecular structures and the impact of mutations on protein structures. The curriculum begins by having students perform Meet the Gene Machine, a play developed by the Science Communication Unit (www.science.uwe.ac.uk/sciencecommunication) at the University of the West of England, funded by the Wellcome Trust, and used with permission (Table 1). The play sets the stage for the rest of the curriculum, helping students explore some of the myths and realities of genetic testing today as they follow the story of a family considering using genetic testing to learn if they possess mutations in BRCA1. Students are also introduced to principles-based bioethics in order to support their thoughtful consideration of the many social and ethical implications of genetic testing. With the Bio-ITEST program’s emphasis on promoting student interest in STEM careers, each lesson features an individual who uses bioinformatics in his or her work or whose work is made possible by bioinformatics (Table 1). In the culminating career lesson (lesson 7), students explore each featured career in greater depth, reading transcripts of interviews with the career professionals and writing their own résumés to document their experience in molecular biology and bioinformatics.

Advanced Curriculum: Using Bioinformatics: Genetic Research

DNA barcoding is a taxonomic method that uses a short genetic marker in an organism’s DNA to identify it as belonging to a particular species (Folmer et al., 1994; Hebert et al., 2003). For animals, the mitochondrial-encoded cytochrome c oxidase subunit 1 (COI) gene is used (Folmer et al., 1994). The advanced curriculum, Using Bioinformatics: Genetic Research, explores DNA barcoding of animals, building on lessons from the introductory curriculum and incorporating additional bioinformatics resources to teach concepts related to species diversity and evolution. DNA barcoding is a technique with several advantages for use in an educational setting. Because the COI gene is mitochondrial, the DNA is more abundant and less prone to degradation (for classes performing their own wet-lab experiments). The region that is sequenced is short, eliminating the need to generate several overlapping sequences and assemble them. Mitochondrial DNA lacks introns in many organisms, which also simplifies the analyses. In addition, DNA barcoding provides concrete connections for students between DNA sequences and the surrounding world. Barcoding can be used to catalogue and confirm the discovery of a new species or to identify the species of an unknown sample. In the Using Bioinformatics: Genetic Research curriculum, students use BLAST to identify an unknown DNA sequence, perform multiple sequence alignments, and build phylogenetic trees (Table 2). They also learn to use the bioinformatics tool ORFinder to identify open reading frames in a DNA sequence.

As in the introductory curriculum, each advanced lesson features a professional who uses bioinformatics in his or her work or whose work is made possible by bioinformatics (Table 2). In the culminating career lesson (lesson 8), students read interview transcripts and perform career-related Internet research. They also build on the career skills developed in Using Bioinformatics: Genetic Testing lesson 7 by updating their résumés, learning to write a cover letter, and participating in mock job interviews.

A key element of the Bio-ITEST program is the incorporation of an authentic bioinformatics research project into the curricular materials, as this has been shown to increase student engagement and interest in STEM careers (O’Neill and Calabrese Barton, 2005). DNA barcoding offers exciting opportunities for students to participate in authentic research, generate testable hypotheses, and learn how to use the tools of science. Lesson 9 is an optional extension lesson in which students learn how to analyze DNA sequence data provided by the Bio-ITEST program or generated in their classrooms (wet lab).

Bioinformatics Teacher Professional Development

NWABR’s professional development workshops are based on five principles of professional development consistent with research on adult learning (Sparks and Hirsh, 1997). These principles include building upon the teacher’s current knowledge and skills and providing engaging and varied opportunities to practice new skills and receive feedback about progress. Successful professional development should result in measurable increases in teacher knowledge and skills that are linked to outcomes in student achievement.

The Bio-ITEST curriculum is shared with teachers in two different professional development formats: a 1.5-d workshop, An Introduction to Bioinformatics, held in late Winter, and a 2-wk workshop, Using Bioinformatics: Genetic Research, held each summer. Both formats provide teachers with the opportunity to see where bioinformatics is used in real-world situations through visits to Seattle-area research institutions and interactions with scientists, and to experience the Bio-ITEST curriculum firsthand. To measure the effect of the Bio-ITEST curriculum and professional development on STEM career awareness and related outcomes, we recruited the 24 teachers from around the country who participated in the 2010 Summer workshop to participate in the Bio-ITEST research study. After teachers completed the workshop, their students were also recruited to participate in the study during the 2010–2011 academic year.

The 2-wk Bio-ITEST Summer workshop, held at Shoreline Community College (SCC) in Shoreline, Washington, exposed teachers to both the introductory and advanced bioinformatics curricula. Given the complex nature of the subject matter and the amount of information covered during only 10 d of instruction, the participating teachers received background readings and homework assignments introducing them to bioinformatics, genetic testing, and DNA barcoding in advance of the workshop. NWABR’s previous experience in professional development has been in the context of instructing teachers on using ethics in their classrooms.
<table>
<thead>
<tr>
<th>Title</th>
<th>Learning objectives</th>
<th>Activities</th>
<th>Featured career and rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lesson 1: Bioinformatics and Genetic Testing</strong></td>
<td>Genetic tests are available for many conditions, but vary in their clinical validity and usefulness. Genetic tests can have social and ethical implications.</td>
<td>Student-led play, <em>Meet the Gene Machine</em> Teacher-led exploration of DTC company website, 23andMe NOVA video, “A Family Disease”</td>
<td>Bioengineer: develops devices like the “gene machine” featured in the play</td>
</tr>
<tr>
<td><strong>Lesson 2: Navigating the NCBI</strong></td>
<td>Biological data are stored in public databases such as the one at the NCBI. Genetic tests are developed using the biological information available in databases such as the one at the NCBI. All organisms need DNA repair proteins like BRCA1, including cats and dogs.</td>
<td>Student-led exploration of the NCBI Understanding databases through a comparison of the NCBI and iTunes</td>
<td>Veterinarian: genetic testing is now available for animals, too</td>
</tr>
<tr>
<td><strong>Lesson 3: Exploring Genetic Testing: A Case Study</strong></td>
<td>Genetic testing can have implications for family members of the patient, as they share the same genetic material. Ethical principles can be applied to many situations, assist in considering alternative perspectives, and facilitate engagement in discussion and decision making.</td>
<td>Structured academic controversy using a short case study about a woman considering BRCA1 genetic testing</td>
<td>Genetic counselor: helps people consider the risks and benefits of genetic testing</td>
</tr>
<tr>
<td><strong>Lesson 4: Understanding Genetic Tests to Detect BRCA1 Mutations</strong></td>
<td>Reference sequences are used to determine whether patient DNA sequences contain mutations. The bioinformatics tool BLAST can be used to compare DNA and protein sequences.</td>
<td>Use a pedigree and Punnett squares to identify family members who should consider testing for BRCA1 mutations Align patient DNA and protein sequences against a reference sequence to identify a mutation using BLAST</td>
<td>Laboratory technician: processes patient samples for genetic testing</td>
</tr>
<tr>
<td><strong>Lesson 5: Learning to Use Cn3D: A Bioinformatics Tool</strong></td>
<td>Bioinformatics tools like Cn3D help scientists visualize molecular structures. A protein is a physical “thing” with a three-dimensional structure that determines its function. A mutation can impact the three-dimensional structure (and therefore the function) of a protein.</td>
<td>Student-led exploration of macromolecular structure using Cn3D Teacher-led exploration of the impact of mutations on the BRCA1 protein using Cn3D</td>
<td>Three-dimensional animator: utilizes biological information to make difficult concepts understandable (such as the animation featured in this lesson)</td>
</tr>
<tr>
<td><strong>Lesson 6: Evaluating Genetic Tests: A Socratic Seminar Discussion</strong></td>
<td>Genetic tests differ in their clinical validity and usefulness. There are some conditions for which there are genetic tests but no effective treatment. Medical conditions differ in their penetrance and the number of genes involved.</td>
<td>Socratic seminar discussion utilizing one of two readings</td>
<td>Bioethicist: helps scientists and society consider the ethical implications of scientific endeavors, including genetic testing</td>
</tr>
<tr>
<td><strong>Lesson 7: An Introduction to Bioinformatics Careers</strong></td>
<td>Bioinformatics tools are used by people in many different careers. Different careers require different skills and education. Jobs in many fields require submission of a résumé specific to that job.</td>
<td>Select a career and read an interview transcript with a career professional from lessons 1–6 Perform Internet research about a selected career Prepare a résumé</td>
<td>Students select one career from previous lessons to explore further</td>
</tr>
<tr>
<td><strong>Lesson 8: Genetic Testing Unit Assessment: ALAD and SOD1</strong></td>
<td>Demonstrate proficiency using BLAST, Cn3D, and ethical reasoning</td>
<td>Application of BLAST, Cn3D, and ethical reasoning skills to a new genetic disease and associated genetic test</td>
<td>None</td>
</tr>
</tbody>
</table>
Table 2. Advanced curriculum, *Using Bioinformatics: Genetic Research*: lesson activities and learning objectives

<table>
<thead>
<tr>
<th>Title</th>
<th>Learning objectives</th>
<th>Activities</th>
<th>Featured career and rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson 1: The Process of Genetic Research</td>
<td>Science is a process involving observations about the natural world, and the generation and testing of hypotheses. Genetic research and bioinformatics can be used to answer research questions in many different STEM fields. DNA sequence data can be used to evaluate species relatedness.</td>
<td>Think–pair–share exploration of genetic research questions in various STEM fields. Review of the scientific process. Student-generated hypothesis of canid relatedness. Pairwise comparisons of short paper DNA sequences.</td>
<td>DNA sequencing core lab manager: helps scientists obtain DNA sequence data for their research studies.</td>
</tr>
<tr>
<td>Lesson 2: DNA Barcoding and the Barcode of Life Database (BOLD)</td>
<td>Scientists often collaborate with one another to conduct research. Biological data are shared by scientists and stored in public databases such as the one at the NCBI and BOLD. The bioinformatics tool BLAST can be used to identify unknown DNA sequences.</td>
<td>Use BLAST to identify “unknown” DNA sequences provided by NWABR. Obtain taxonomic information about species using BOLD. Form collaborative groups with other students whose identified species are in the same taxonomic class.</td>
<td>Postdoctoral scientist in DNA and history: uses genetic data to study the history of human populations and migrations.</td>
</tr>
<tr>
<td>Lesson 3: Using Bioinformatics to Study Evolutionary Relationships</td>
<td>Scientific collaboration and data sharing are vital to the scientific process. Bioinformatics tools like JalView/ClustalW2 can be used to analyze long DNA sequences. Phylogenetic trees can be used to draw conclusions about evolutionary relationships.</td>
<td>Use JalView/ClustalW2 and DNA sequence data from lesson 2 to compare multiple sequences. Use BLAST and an outgroup (provided by NWABR) to create a phylogenetic tree and draw conclusions about evolutionary relationships.</td>
<td>Microbiologist: uses genetic data to study microbes that cause diseases such as tuberculosis or influenza.</td>
</tr>
<tr>
<td>Lesson 4: Using Bioinformatics to Analyze Protein Sequences</td>
<td>DNA is composed of two strands that are complementary and antiparallel. There are six potential reading frames for protein translation in each strand of DNA. Bioinformatics tools can be used to identify open reading frames and compare protein sequences.</td>
<td>Paper exercise to understand the complementary nature of DNA and six reading frames of protein translation. Use ORFinder to identify the likely reading frame for a DNA sequence. Perform multiple sequence alignment using a group’s protein sequences.</td>
<td>Biological anthropologist: uses genetic data to study the evolution of humans and other hominids.</td>
</tr>
<tr>
<td>Lesson 5: Protein Structure and Function: A Molecular Murder Mystery</td>
<td>Mitochondria are the site of ATP production in the cell. Cytochrome c oxidase is involved in ATP production. The active site of a protein is vital to the function of the protein. Substances that bind to the active site can interfere with protein function.</td>
<td>Identify the active site of cytochrome c oxidase using Cn3D. Identify a foreign substance (a poison) bound to the active site of cytochrome oxidase.</td>
<td>Molecular diagnostics researcher: uses genetic information about infectious organisms to develop diagnostic tests.</td>
</tr>
<tr>
<td>Lesson 6: Assessment: Writing Research Reports</td>
<td>Scientists share their work with other scientists in the spirit of collaboration and to advance scientific knowledge. The components of a research report correspond to the steps of the scientific method.</td>
<td>Write a research report with instruction, methods, results, and discussion sections and figures. Assessment alternatives: scientific poster, scientific abstract, or a science-related magazine article.</td>
<td>Science and technical writer: helps scientists communicate effectively to the public and to other scientists.</td>
</tr>
</tbody>
</table>

(Continued).
Table 2. Continued

<table>
<thead>
<tr>
<th>Title</th>
<th>Learning objectives</th>
<th>Activities</th>
<th>Featured career and rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lesson 7: Who Should Pay? Funding Research on Rare Genetic Diseases</td>
<td>Rare genetic conditions affect a limited number of people but can cause great suffering. Much scientific research in the United States is funded by taxpayer money. There is a limited amount of money that must be allocated based on our values and the needs of stakeholders. Bioethical principles can provide a structure for making complex decisions.</td>
<td>Jigsaw exercise: meet in “like” and then “mixed” groups of stakeholders (parent, researcher, doctor, or advocate) Use bioethical principles to draft recommendations on allocation of public resources for research on rare genetic diseases</td>
<td>Pediatric neurologist: uses genetic testing results to help diagnose and treat children with diseases of the brain or spinal cord</td>
</tr>
<tr>
<td>Lesson 8: Exploring Bioinformatics Careers</td>
<td>Bioinformatics tools are used by people in many different careers. Different careers require different skills and education. Jobs in many fields require submission of a résumé and cover letter specific to that job. Job interviews include questions about your skills and experience (optional).</td>
<td>Select a career and read an interview transcript with a career professional from lessons 1–6 Perform Internet research about a selected career Create or update a résumé Critique and write a cover letter Mock job interview (optional)</td>
<td>Students select one career from previous lessons to explore further</td>
</tr>
<tr>
<td>Lesson 9: Analyzing DNA Sequences and DNA Barcoding</td>
<td>DNA sequences can be used to identify the origin of samples. DNA data (called a chromatogram) are generated by DNA sequencing. For increased accuracy, both strands of DNA are often sequenced. Data can be used to guide decision-making when reconstructing a DNA sequence.</td>
<td>Use BLAST and FinchTV to analyze DNA chromatograms (provided by NWABR or generated in class using the wet lab) Identify and edit discrepancies between sequence data from both strands of DNA Use a phylogenetic tree from BOLD for sample identification</td>
<td>None</td>
</tr>
<tr>
<td>Wet lab: DNA Barcoding: From Samples to Sequences</td>
<td>DNA barcoding involves multiple laboratory experiments prior to bioinformatics analysis. DNA must be purified through a process involving cell lysis and separation of the DNA from the rest of the cell debris. PCR is used to make many copies of a gene or region for use in subsequent analyses. Agarose gel electrophoresis is performed to confirm whether a PCR was successful. A purified DNA product is used for DNA sequencing.</td>
<td>Lab 1: DNA purification for DNA barcoding Lab 2: Copying the DNA barcoding gene using PCR Lab 3: Analyzing PCR results with agarose gel electrophoresis Lab 4: Preparation of PCR samples for DNA sequencing</td>
<td>None</td>
</tr>
</tbody>
</table>

(Wishes), to identify topics that needed additional clarification, and to modify instruction for the next day. Teachers received training in using the introductory lesson materials, including review of pedagogical strategies included in the procedures section of each lesson, and experienced each of the full-length lessons themselves in order to practice their new skills.

As the focus of the advanced curriculum is DNA barcoding, a primary goal of the workshop was for teachers to experience the entire barcoding process: obtaining a sample from an organism, purifying the DNA, using polymerase chain reaction (PCR) to amplify the barcode sequence, sending it to be sequenced, analyzing the DNA data, and comparing those data with sequences from database repositories at the NCBI (www.ncbi.nlm.nih.gov) and the Barcode of Life Data Systems (http://barcodinglife.com). Teachers isolated DNA from an “unknown” tissue sample (Oncorhynchus kisutch [Coho salmon] purchased at a local grocery store), used PCR, checked their PCR products by agarose gel electrophoresis, purified their PCR products, and submitted them for sequencing (see “wet lab” in Table 2). Experimental protocols from commercially available kits and protocols obtained from members of the DNA barcoding community (Yancy et al., 2009) were adapted for classroom use. These
protocols were supplemented with educational support resources available online from the Howard Hughes Medical Institute and the DNA Learning Center (www.dnalc.org and www.dnai.org). Teachers used the DNA sequence data they generated in week 1 for their exploration of the advanced strand lessons during week 2 to help them become more familiar with using bioinformatics tools and databases. Teachers also received training in using software (FinchTV; Geospiza, Seattle, WA) and online programs for analyzing DNA sequences, performing multiple sequence alignments, and using bioinformatics to study evolutionary relationships. The goal of the activities was to help teachers understand the flow of biological data from the lab bench to the computer.

Echoing the 2009 curriculum development workshop format, Summer workshop teachers deepened their exploration of bioinformatics and related careers by touring local research facilities and learning about next-generation DNA-sequencing technology and other high-throughput data-generation and analysis techniques. Guest speakers and panel discussions with scientists who perform genetic research were included in the 2-wk program to illustrate diverse careers and areas of research influenced by bioinformatics.

### Bio-ITEST Program Evaluation and Research Study

An external evaluation team conducted a formative and summative program evaluation that addressed two questions:

1. In what ways does the Bio-ITEST model of curriculum development and teacher professional development add to our understanding of how to best prepare teachers to develop the knowledge and skills necessary among their students for participation in the STEM workforce?
2. In what ways does the Bio-ITEST contribute to our understanding of how to engender student awareness of and interest in STEM careers? The program evaluation used teacher and student structured interviews, focus groups, site observations of professional development and classroom activities, and numerical and open-ended survey questions to determine program effectiveness.

As part of this larger program evaluation, the Bio-ITEST research study focused on the following questions:

- What were the effects of Bio-ITEST program participation on teachers’ knowledge and perceptions of bioinformatics and related STEM careers?
- What were the effects of Bio-ITEST participation on students’ knowledge and perceptions of bioinformatics and related STEM careers?
- Did change in participating teachers’ knowledge and perceptions correlate with change in students’ knowledge and perceptions of bioinformatics and related STEM careers?

All 24 teacher participants from the 2010 Summer Bio-ITEST professional development workshop Using Bioinformatics: Genetic Research were recruited to take part in the evaluation and the research study. The Bio-ITEST research study utilized pre- and postsurveys of teacher and student participants to measure changes in awareness of STEM career opportunities, particularly in bioinformatics (“awareness”); sense of self-efficacy using bioinformatics tools (“self-efficacy”); perceptions of the relevance of biology content to their lives (“relevance”); and engagement with science (“engagement”). Teacher results are based on surveys tracking growth over three points in time: at the beginning (baseline, “preworkshop”) and end (“postworkshop”) of the August 2010 Using Bioinformatics: Genetic Research workshop, and at the end of the school year (May or June 2011, “end of year”). An additional goal of the study was to determine whether changes in teachers’ perceptions correlated with changes in students’ perceptions in each of the four study areas (awareness, self-efficacy, relevance, and engagement). Two hundred eighty-nine students of participating teachers who taught the introductory unit, Using Bioinformatics: Genetic Testing, completed pre- and postunit surveys. Surveys were administered online where possible or via paper and pencil. All teacher and student surveys are available in the Supplemental Material.

In addition to the correlational research study, the program evaluation addressed the question of student impacts by soliciting feedback from teachers and students through open-ended queries on three questionnaires: a prequestionnaire for the May 2011 reunion of Bio-ITEST teachers, the Bio-ITEST Educators End of Year Survey, and the Bio-ITEST (advanced unit) Genetic Research Curriculum Student Survey. Teacher comments are based on their experiences teaching both the introductory and advanced curriculum units and focused on the following themes: impacts of the professional development workshop, impacts of the curriculum units and program overall, and effectiveness of conveying bioinformatics and related career materials to students.

### Institutional Review Board (IRB) Approval

This study was reviewed and approved by Quorum Review IRB (Quorum Review File # 24134/1). All study participants and/or their legal guardians provided written informed consent for the collection and subsequent analysis of verbal and written responses.

### Survey Development

Face validity for the survey pre/postunit constructs (awareness, self-efficacy, engagement, as well as relevance for students) was established through an iterative item-construction process by the research team, and the content validity of the four constructs was empirically tested using exploratory factor analysis (EFA) of the preunit survey items (Stevens, 2002; Tabachnick and Fidell, 2007). The EFA used maximum-likelihood estimation (Tabachnick and Fidell, 2007) and a Varimax orthogonal rotation. For the teacher survey, an EFA showed that the three-factor model fitted the data well ($\chi^2(5) = 0.84, p > 0.05$), and the set of factors accounted for 68% of the variance in the items (internal consistency across all items was 0.84). For student results, an EFA showed that a four-factor solution fitted the data fairly well, in that all communalities were $> 0.40$, and further, the set of four factors together accounted for 61% of the total variance in the set of items. The results of the four-factor EFA solution show that the item-factor loadings corresponded well with the four constructs overall. Internal consistencies were as follows: awareness: 0.85; relevance: 0.65; self-efficacy: 0.76; and engagement: 0.83.
Table 3. Characteristics of teacher participantsa

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>All 2010 teacher participants</th>
<th>Curriculum implementers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of teachers</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td>Gender</td>
<td>75% Female (18)</td>
<td>67% Female (8)</td>
</tr>
<tr>
<td></td>
<td>25% Male (6)</td>
<td>33% Male (4)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>75% Non-Hispanic white (18)</td>
<td>75% Non-Hispanic white (9)</td>
</tr>
<tr>
<td></td>
<td>4% Hispanic (1)</td>
<td>0% Hispanic (0)</td>
</tr>
<tr>
<td></td>
<td>21% Unknown (5)</td>
<td>25% Unknown (3)</td>
</tr>
<tr>
<td>Race</td>
<td>75% White (18)</td>
<td>67% White (8)</td>
</tr>
<tr>
<td></td>
<td>4% Black/African American (1)</td>
<td>0% Black/African American (0)</td>
</tr>
<tr>
<td></td>
<td>8% Asian/Southeast Asian (2)</td>
<td>17% Asian/Southeast Asian (2)</td>
</tr>
<tr>
<td></td>
<td>0% American Indian (0)</td>
<td>0% American Indian (0)</td>
</tr>
<tr>
<td></td>
<td>0% Alaska Native (0)</td>
<td>0% Alaska Native (0)</td>
</tr>
<tr>
<td></td>
<td>0% Native Hawaiian (0)</td>
<td>0% Native Hawaiian (0)</td>
</tr>
<tr>
<td></td>
<td>0% Pacific Islander (0)</td>
<td>0% Pacific Islander (0)</td>
</tr>
<tr>
<td></td>
<td>8% Other (2)</td>
<td>17% Other (2)</td>
</tr>
<tr>
<td>Highest level of education completed</td>
<td>8% Doctorate (2)</td>
<td>8% Doctorate (1)</td>
</tr>
<tr>
<td></td>
<td>79% Master’s degree (19)</td>
<td>83% Master’s degree (10)</td>
</tr>
<tr>
<td></td>
<td>13% Bachelor’s degree (3)</td>
<td>8% Bachelor’s degree (1)</td>
</tr>
<tr>
<td>Certifications</td>
<td>83% Biology (20)</td>
<td>83% Biology (10)</td>
</tr>
<tr>
<td></td>
<td>63% Other science (15)</td>
<td>50% Other science (6)</td>
</tr>
<tr>
<td></td>
<td>13% CTEb (3)</td>
<td>17% CTEb (2)</td>
</tr>
<tr>
<td></td>
<td>13% Conditional (3)</td>
<td>0% Conditional (0)</td>
</tr>
<tr>
<td>Prior professional development</td>
<td>63% Ethics (15)</td>
<td>58% Ethics (7)</td>
</tr>
<tr>
<td></td>
<td>33% Bioinformatics (8)</td>
<td>50% Bioinformatics (6)</td>
</tr>
<tr>
<td>Mean years of teaching experience</td>
<td>13 High school</td>
<td>13 High school</td>
</tr>
<tr>
<td></td>
<td>10 Biology</td>
<td>12 Biology</td>
</tr>
<tr>
<td></td>
<td>13 All sciencesc</td>
<td>14 All sciencesc</td>
</tr>
</tbody>
</table>

aPercentages of individual items may not total 100% due to rounding or classification of individuals into multiple categories.
bCareer and Technical Education.
cIncludes biology.

Statistical Analyses

Basic one-sample t tests were used to evaluate whether teachers’ pre- to postworkshop and postworkshop to end-year changes were statistically significant, as well as whether students’ pre/postunit changes were statistically significant. We corrected for type I error inflation familywise by grouping the statistical tests by survey item type (i.e., teachers’ pre-post responses, teachers’ retrospective responses, introductory unit students’ responses, introductory unit students’ retrospective responses, and advanced unit students’ retrospective responses) and adjusting our per-comparison type I error rate using the Bonferroni adjustment (which divides the alpha level by the number of tests performed). Familywise alpha level was set at 0.05. SPSS/PASW (IBM Corp., Armonk, NY) was used for these analyses.

In addition to item-level t tests, the correlations between teacher and student gains were tested within a multilevel modeling framework using HLM 7 (Raudenbush et al., 2011). This method of analysis accounts for the nonindependence of students’ scores within a classroom and uses appropriate degrees of freedom for testing teacher gain correlations with student gains. For these analyses, we created composite construct scores for each teacher and student at each survey wave (pre- and postunit) by computing the mean of the scores of each item related to the construct (see prior discussion of EFA results in Survey Development). Twelve of the 24 workshop teachers participated in this phase of the research. Three retrospective items (administered only on the postunit survey) asking students to estimate their knowledge and skills “before” and “after” the unit, were computed as the students’ difference scores.

RESULTS

Teacher Effects: Impacts of the Professional Development Experience

Background Characteristics. Characteristics of the 24 teachers who participated in the 2010 Summer professional development workshop, Using Bioinformatics: Genetic Research, are summarized in Table 3. Three-quarters of workshop participants were female and white, and of those indicating ethnicity, one teacher was Hispanic/Latino. Teachers represented 21 different schools. Most had master’s degrees (79%), as well as teaching certifications in biology (83%) and/or another science certification (63%). Three teachers (12%) had career or technical education (CTE) certifications. This group of teachers was quite experienced overall, with an average of nearly 13 yr of high school teaching experience and over 10 yr spent teaching biology. One-third of the teachers had prior professional development in bioinformatics (including two who participated in the 2010 1.5-d workshop, An
Introduction to Bioinformatics, and one who participated in the 2009 Bio-ITEST curriculum development workshop. Nearly two-thirds (63%) had prior professional development in integrating ethics into science curricula.

Qualitative Findings. As part of the postworkshop survey, teachers had the opportunity to comment on the “most significant take-aways” from their workshop experiences. These remarks were supplemented by the teacher interviews and focus groups conducted by the evaluation team. Many participant comments related to the general theme of increased understanding of bioinformatics, biology, and biotechnology, and greater comfort in the skills they had learned:

“I am so excited that I am gaining some comfort and familiarity with bioinformatics software and databases; it will become part of the tool kit I can use when designing lessons.”

“An additional take-away is a much more well-formed understanding of how bioinformatics fits with biological research and biotechnology. I had a very unclear understanding of this prior to the workshop, and now feel I can easily articulate this understanding to my students, particularly those that show interest and aptitude both in biology and technology.”

Teachers also remarked on the wealth of bioinformatics resources available to teachers and the importance of hands-on experience using them, including the curriculum materials, bioinformatics databases, and professionals in the field who are interested in providing support:

“I feel confident that I’ll be able to turn around and teach it to my students this next school year. The curriculum is extremely well thought out, very teacher friendly, and will be interesting to students. I cannot wait to bring this curriculum to my school, and I am very proud that I’m able to offer my students an opportunity to learn about and actually do science that is on the leading edge of what is being done in our area.”

In addition, teachers noted the importance of exposing students to STEM careers, bioinformatics, and bioethics:

“Two things that I’ve been trying to bring to my classes is [sic] already woven into the curriculum: ethical studies, and career information. Before coming to this class, I realized that I’ve done a good job of getting students interested in science, but a poor job of guiding them toward careers based on that interest.”

“The most significant take-away from this workshop is the importance of exposing students to STEM careers, particularly those using bioinformatics. This is an area in which I have much room for growth and really value the resources provided.”

Teacher Survey Findings. The effects of the 2010 Bio-ITEST Summer professional development workshop Using Bioinformatics: Genetic Research were evident in survey findings measuring teacher self-reported changes in career awareness, engagement and self-efficacy at three points in time (Figure 1): before the Summer workshop (“pre-workshop”), immediately after the Summer workshop (“postworkshop”), and at the end of the school year (“end of year”). When compared with responses to survey questions on the preworkshop survey (n = 24), teachers showed significant increases postworkshop (n = 24) and at the end of the academic year (n = 23) in all three conceptual areas measured. All pre/postworkshop gains were statistically significant (adjusted for multiple comparisons using the Bonferroni procedure), with the largest gains in the areas of career awareness and self-efficacy using the tools of bioinformatics. All preworkshop/end-of-year gains were statistically significant, except “My interest in analyzing biological information,” a measure of engagement. Mean ratings generally declined somewhat from postworkshop to end of year, but declines on three of the items were not statistically significant, and all end-of-year ratings were still significantly higher than the corresponding preworkshop means.

Retrospective questions on the postworkshop and end-of-year surveys asked teachers to rate themselves “before” the workshop (retrospectively) and “now” (at the end of the workshop) on survey items measuring career awareness and self-efficacy. Similar retrospective questions were asked of teachers on the end-of-year survey. Teachers demonstrated statistically significant gains on all retrospective survey items (after adjusting for multiple comparisons) following the 2010 professional development workshop “Using Bioinformatics: Genetic Research” (n = 24) and at the end of the 2010–2011 academic year (n = 23; Figure 2), with the largest reported gains for self-efficacy items.

Student Effects: Impact of the Introductory Curriculum Unit

Of the 24 teachers who participated in the 2010 Summer workshop, 13 are known to have implemented four or more of the eight introductory lessons, of whom 12 returned consent forms to allow inclusion of their students’ data in...
analyses. Chi-square and t tests were used to compare the demographic characteristics shown in Table 3 of all 2010 teacher participants and the 12 curriculum implementers who participated in the research study. No significant differences were found.

The introductory lessons were trialed in at least 28 classes, with students representing various levels of science background and interest, and in a diversity of settings. For example, although the lessons are designed primarily for classroom use, one of the teachers introduced both the introductory and advanced lessons in an after school club. At least two of the teachers implemented the lessons in courses that were part of a science career track aimed at highly motivated science students. In other cases, the lessons were introduced in a required biology class, in which students may not have much prior interest in science. Of the 699 students in the 28 classrooms of the 12 teachers included in the research study, 374 (54%) students consented to take part in the Bio-ITEST study and 289 (41%) completed both preunit and postunit surveys measuring conceptual constructs similar to those measured for their teachers. Characteristics of student participants are shown in Table 4.

Students using the introductory curriculum made significant pre/postunit gains (adjusted for multiple comparisons) on all items measuring awareness and self-efficacy (Figure 3). The pre/postunit changes on the two relevance items were not significant. One relevance item, “I am interested in how science knowledge can guide ethical decision making,” showed little change preunit (mean = 5.19) to postunit (mean = 5.06). The second relevance item on the pre/postunit surveys, “I think it is important for people in our society to learn about science,” showed high preunit scores (mean = 5.76) that changed little postunit (mean = 5.86). However, students reported significant gains on the retrospective postunit item measuring relevance, evaluating their perceptions of the connection between biology content and their personal lives (“before this unit” mean = 3.88; “now” mean = 5.64; Figure 4). The retrospective survey items

<table>
<thead>
<tr>
<th>Curriculum unit</th>
<th>Introductory</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of students</td>
<td>289</td>
<td>41</td>
</tr>
<tr>
<td>Gender</td>
<td>63%</td>
<td>56%</td>
</tr>
<tr>
<td>37%</td>
<td>Female (181)</td>
<td>Female (23)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>95%</td>
<td>90%</td>
</tr>
<tr>
<td>4%</td>
<td>Non-Hispanic white (274)</td>
<td>Non-Hispanic white (37)</td>
</tr>
<tr>
<td>1%</td>
<td>Hispanic (13)</td>
<td>Hispanic (2)</td>
</tr>
<tr>
<td>Race</td>
<td>70%</td>
<td>68%</td>
</tr>
<tr>
<td>6%</td>
<td>White (203)</td>
<td>White (28)</td>
</tr>
<tr>
<td>13%</td>
<td>Black/African American (17)</td>
<td>12%</td>
</tr>
<tr>
<td>1%</td>
<td>Asian/Southeast Asian (37)</td>
<td>10%</td>
</tr>
<tr>
<td>0%</td>
<td>American Indian (3)</td>
<td>2%</td>
</tr>
<tr>
<td>0%</td>
<td>Alaska Native (0)</td>
<td>0%</td>
</tr>
<tr>
<td>1%</td>
<td>Native Hawaiian (2)</td>
<td>2%</td>
</tr>
<tr>
<td>1%</td>
<td>Pacific Islander (3)</td>
<td>0%</td>
</tr>
<tr>
<td>8%</td>
<td>Other (23)</td>
<td>0%</td>
</tr>
<tr>
<td>0%</td>
<td>Unknown (0)</td>
<td>5%</td>
</tr>
<tr>
<td>Grade level</td>
<td>18%</td>
<td>0%</td>
</tr>
<tr>
<td>33%</td>
<td>Freshman (51)</td>
<td>Freshman (0)</td>
</tr>
<tr>
<td>25%</td>
<td>Sophomore (95)</td>
<td>2%</td>
</tr>
<tr>
<td>24%</td>
<td>Junior (73)</td>
<td>32%</td>
</tr>
<tr>
<td>24%</td>
<td>Senior (69)</td>
<td>63%</td>
</tr>
<tr>
<td>0%</td>
<td>Unknown (0)</td>
<td>2%</td>
</tr>
</tbody>
</table>

*Percentages of individual items may not total 100% due to rounding.
measuring self-efficacy and awareness also showed significant gains (greater than two points).

Two of the four pre/postunit survey items measuring engagement showed no significant change. Students began the unit with a fairly high level of engagement, indicated by a mean of 5.34 on the question “In general, I enjoy learning about science topics.” Postunit, this score remained essentially unchanged (mean = 5.37). Student preunit scores were similar, though slightly lower (mean = 4.88), on a survey item measuring engagement with STEM careers (“I see myself working in a career that involves scientific information.”). These scores also changed little postunit (mean = 4.90). Two of the four pre/postunit changes on the engagement items were significant and negative. One question measured engagement with computer programs to visualize three-dimensional images of molecules (i.e., Cn3D), while the other queried about interest in analyzing biological information. These declines were modest, < 0.5 on the 7-point scale (Figure 3). The retrospective postunit survey did not measure student engagement.

To further explore student engagement and the relationship between engagement and self-efficacy, we performed correlational analyses on the pre/postunit responses of the 289 students who participated in the introductory lessons. The preunit correlation between engagement and self-efficacy was \( r = 0.50 \) (25% shared variability; \( p < 0.001 \)), and the postunit correlation increased to \( r = 0.63 \) (40% shared variability). The correlation between pre–post gains on engagement and pre–post gains on self-efficacy was \( r = 0.34 \) (12% shared variance).

**Student Effects: Impact of the Advanced Curriculum Unit**

A postunit survey measured the students’ perceived effects from the advanced genetic research curriculum. Three teachers who implemented the unit in three classrooms were included in this study. These three teachers were more experienced teaching at the high school level (mean = 23 yr) than the nine teachers who only implemented the introductory unit (mean = 9 yr). These three teachers also scored higher on preworkshop survey measures of self-efficacy (mean = 3.67 vs. mean = 5.00). However, these differences did not remain statistically significant after Bonferroni adjustment for multiple comparisons. Usable surveys were obtained from a total of 41 advanced unit students. Characteristics of student participants are shown in Table 3.

Student postunit retrospective mean ratings comparing “before this unit” with “now” showed large gains on all items, ranging from 1.1 to 3.1 on the 7-point scale (Figure 5). The largest increases were among items measuring self-efficacy (confidence in accessing biological databases and understanding how databases store biological information) and engagement (interest in analyzing biological information). Additional survey items measuring engagement (see myself working in a STEM career and interest in using computer programs to visualize three-dimensional images of molecules) also showed gains of 1.3–1.9 points. All changes were significant, even after adjusting for multiple comparisons.

**Teacher and Student Comments on Student Effects**

Teachers and students were asked to reflect on their most significant take-aways or lessons learned from the bioinformatics curriculum units. Teacher comments are based on their experiences teaching both the introductory and advanced curriculum units. Student comments generally represent more capable students, many of whom were already interested in science, because only the advanced curriculum survey collected open-ended student responses. Comments were categorized according to each of the four career constructs. Some quotes could be classified into more than one category.
but are assigned to one construct for simplicity. A representative sample of comments is presented in Table 5.

**Correlation of Teacher and Student Gains**

We tested whether or not participating teachers’ pre/postworkshop gains in knowledge and perceptions of bioinformatics and related STEM careers correlated with growth in students’ knowledge and perceptions of bioinformatics and related STEM careers using multilevel modeling (students [level 1] nested within teachers [level 2]). Results showed no evidence of a relationship between teacher and student growth on any construct, including teacher’s total pre/postworkshop change, total postworkshop (point in time) composite scores, and total end-of-year (point in time) composite scores, as well as postworkshop and end-of-year retrospective ratings. In other words, teacher subscale composites did not correlate with student composites (i.e., teacher career awareness did not correlate with student career awareness and vice versa). It is important, however, to note that these analyses were severely limited by a small teacher sample size ($n = 12$ available) and a lack of data to control for lesson fidelity implementation effects on student gains.

**DISCUSSION**

The Bio-ITEST program responds directly to the learning goal of the NSF’s strategic plan: to cultivate a world-class and broadly inclusive science and engineering workforce and to expand the scientific literacy of all citizens (NSF, 2006). Our program emphasizes cutting-edge bioinformatics resources and provides a broad range of instructional materials that serve both highly motivated science students and the broad base of general biology students who will comprise the majority of our future citizenry. NWABR’s teacher workshops are based on five principles of professional development consistent with research on adult learning (Sparks and Hirsch, 1997). Successful professional development experiences build upon the learner’s current knowledge, provide multiple opportunities to practice new skills, provide ongoing feedback on the learner’s performance, and are linked to measurable outcomes in student performance. Utilizing this model, NWABR’s professional development workshops included teachers as program partners, valuing their prior experiences and building on their existing knowledge during training and curriculum development. All teacher workshops offered extensive opportunities to practice new skills and to give feedback on both teacher and staff performance. Teachers who participated in the Bio-ITEST program had significant increases in awareness of bioinformatics and related STEM careers, as well as improved reports of self-efficacy using bioinformatics tools and databases and integrating these resources into their classrooms. On the basis of numerical and open-ended responses, following instruction with the Bio-ITEST curriculum units, students demonstrated increased understanding of the application of information technologies to the biological sciences, the ethical implications of the acquisition and use of biological information, and the career possibilities in the fields of bioinformatics and related careers. While the Bio-ITEST research study did not find a correlation between changes in teacher and student knowledge and perceptions of STEM careers, the magnitude of change among both groups was significant given the short duration of the educational intervention. In the following sections we highlight some of our key program findings in the context of suggestions for teacher professional development and student instruction in other STEM fields.

**Lessons for Effective Teacher Professional Development**

**Collaborate with Teachers as Program Partners.** Working directly with teachers as respected and valued partners throughout the iterative curriculum-development process, including during lesson design, field-testing, and revision, is a crucial component for program success. In addition to being familiar with state and national education standards, teachers are uniquely qualified to identify instructional approaches and lesson components that will appeal to today’s students. Involving teachers with every phase of curriculum development provides ample opportunity for both oral and written lesson feedback and revision, including identification of areas in which teachers or students may struggle with material in the classroom. Feedback can be solicited from teachers as they experience the lessons in the professional development workshops and following implementation of the lessons with their students in the classroom. For our bioinformatics curricula, teacher feedback led to the expansion of teacher background and procedure instructions in each lesson, as well as incorporation of multiple screen capture images in student instruction handouts to help students navigate and use bioinformatics tools and databases. Improved visual aids permitted most lessons to be completed in the time allotted to the average high school class and reduced reports of student frustration. The modifications teachers made when implementing lessons were incorporated into later versions—most notably in the case of career activities, such as attending a “social mixer,” participating in mock job interviews, and writing scientific abstracts and bioinformatics magazine articles. During professional development workshops, Stars and Wishes forms,
group discussions, and opportunities for teachers to critique curriculum lessons provided ongoing program feedback and assessment and helped to foster a sense of community among teachers and program staff. Respecting teachers as peers and active participants in the iterative curriculum-development process is likely to increase their sense of ownership of the curricular materials, as well as improve the likelihood of successfully meeting program goals.

**Support Teachers as Students.** Following the 2-wk professional development for teachers in Summer 2010, teachers demonstrated significant gains in the areas of career awareness, self-efficacy, and engagement, and these gains were largely sustained throughout the 2010–2011 academic year. Comments from teachers revealed numerous benefits of the professional development model utilized by NWABR. Teachers “became the students,” experiencing each lesson “first-hand,” as their students would. This provided opportunities to ask questions and gain insights about how each lesson could be implemented in their classrooms, to learn how advances in technology had contributed to the materials presented, and to make more explicit connections between curriculum content and STEM careers. Teachers also noted that hands-on wet labs and computer activities provided many opportunities for them to become more familiar with computational and bioinformatics tools. This is supported by the

| Table 5. Teacher and student comments on student effects |
|----------------|---------------|----------------|
| **Construct** | **Teacher comments** | **Student comments** |
| Awareness | “It opened up a whole world to them—they knew nothing about the topic before. Now they understand and can use some bioinformatics tools, and they have a clear understanding that there are jobs available in this area, as well as some knowledge about the types of jobs, and the education required to get them.” | “Some of the most important things were learning about different careers in this unit.” |
| | “They also realized that there are SO MANY career opportunities they had never heard of.” | “I learned that bioinformatics is extremely useful in a wide variety of careers and applications.” |
| | “Introduced the students to careers they had not thought about before. Infusing career-awareness into my curriculum has not been something I have really done before, so the bioinformatics unit was really the only exposure the students had all year.” | “It really introduced me to new possible career choices.” |
| Relevance | “The lessons on ethical issues and awareness of the different careers that use bioinformatics had the most impact. Understanding how technology has changed science and how many different career options there are in biology. I have had many students tell me ‘I didn’t know I could do this in science’—it really is an open ended, making connections ‘real’ curriculum.” | “It opened a door that I could go through, it introduced me into something I might be interested in.” |
| | “They were excited about NCBI. They are so good on computers anyway, one kid became the class teacher. He helped everyone else. Some of them were surprised they could use a tool like that even though they are not scientists.” | “I still want to be a mechanical or electrical engineer, but I might be interested in designing systems to work with biological and bioinformatics technologies.” |
| | “They gained significant confidence in their ability to read, understand, and analyze data. It was fabulous!” | “I already wanted to pursue a career in the biomedical field, so this unit just added interest to the field.” “There are many ways to help people besides being a doctor.” |
| | “The ability to use various tools and databases increased the students’ skills and confidence in applying biology topics.” | “There are different sites available for the public to use for themselves than relying on others to do it for them, and it allows others to learn how the whole process works.” |
| Engagement | “Students told me that they enjoyed this unit . . . It really opened their eyes to new ideas and scientific ways of thinking.” | “I didn’t know it was so easy to access that information.” |
| | “They were really into the Gene Machine and role playing . . . They were also interested in the recurring theme of the breast cancer gene and how it could be looked at from multiple levels of understanding, for example the consequences of the disease and the molecular structure of a mutated protein.” | “That there are massive databases online that you can plug DNA into and get results of which species it is.” |
| | “They were fascinated by 23andMe. They really used it. I took it one step further, and did a genetics project. I had them look for diseases in their families on the website (e.g., sickle cell anemia), and they did a research project which they presented with PowerPoint. They learned more by them doing the research themselves.” | “I am very interested in a way I can work in bioinformatics and combine that with engineering and physics.” |
| | “[The] DNA barcoding unit made me consider more database related careers in science. Before I considered science careers using primarily databases to be boring jobs but now I think it would be very interesting and more than just sitting at a computer all day.” | “[The] DNA barcoding unit made me consider more database related careers in science. Before I considered science careers using primarily databases to be boring jobs but now I think it would be very interesting and more than just sitting at a computer all day.” |
| | “I am now looking at pursuing a career in biological research for the benefit of global health.” | “I am now looking at pursuing a career in biological research for the benefit of global health.” |
| | “It created a new possible job career. I love science and I never knew much about this type of science and it is very fascinating.” | “It created a new possible job career. I love science and I never knew much about this type of science and it is very fascinating.” |
significant retrospective postworkshop gains in teachers’ perceptions of their self-efficacy (Figure 2).

Having teachers experience the curriculum lessons helped uncover a number of teacher misconceptions related to bioinformatics and molecular biology. These included a lack of understanding or knowledge about: the presence of genes on both strands of DNA in a given chromosome; the strand of DNA (anti-sense vs. sense) used for protein transcription and translation; ATG/AUG codons not necessarily acting as “start” codons; proteins’ ability to form bonds to metal ions; and the shape of phylogenetic trees varies with the genes and organisms that are chosen for analysis. These misconceptions were explicitly addressed in later versions of the lessons and teacher professional development workshops. Ample time for teacher feedback, opportunities and diverse venues for questions, and support by program staff and guest scientists are especially valuable when teaching complex subjects such as bioinformatics.

It became clear during the workshops that teachers varied widely in their technical skills. During the Summer 2010 workshop, it became clear that there was a need for instruction in the basic computer skills necessary for bioinformatics analyses, such as downloading files, copying and pasting text (DNA and protein sequences), understanding file formats, finding specific text or sequences within a document or on a Web page, capturing screen images, and bookmarking Web pages. This led to the implementation of Computer Skills 101 in subsequent Summer professional development workshops (i.e., 2011 and 2012) and additional computer instructions in lesson revisions. Many of these skills also proved to be useful for teachers in other settings. Preworkshop surveys of teacher computer skills, practice exercises on preworkshop homework, and having additional program staff and guest scientists available to assist early in the workshops also helped improve instruction. Peer mentoring (pairing more experienced teachers with less experienced teachers during workshop activities) was also an effective approach, resulting in both enhanced learning and a sense of teamwork and camaraderie.

Hands-on experiences using the tools of science, time for questions and uncovering common misconceptions that can present barriers to learning, preassessment of existing skills, and peer mentoring are all effective approaches to teacher professional development that can be implemented in a variety of STEM fields.

Promote Career Awareness Among Teachers by Including Diverse STEM Professionals. Teachers indicated that the chance to network with professional scientists, as well as with other science teachers, was of great benefit to them. Including scientists throughout the lessons— as tour guides, members of panel discussions, and during one-on-one interactions— allowed teachers to interact with and question the scientists about their work and experiences using bioinformatics on a daily basis. Personal stories from scientists, such as how they chose their current career and what they love most (and least) about their jobs as scientists, resonated with teachers and students alike. These anecdotes help humanize scientists for both teachers and students. Exposure to diverse career professionals illustrates the many different approaches that are utilized in a particular field, and the diversity of career professionals themselves may help to dispel many of the stereotypes of scientists that persist in the American psyche.

Lessons for Promoting STEM Careers among Students

Integrate Information about STEM Professionals into Curriculum Units. Introductory unit students made significant pre/postunit gains on survey items measuring career awareness and self-efficacy, as well as on postunit retrospective survey items measuring relevance and self-efficacy. Similar gains were found among advanced unit students. Intentional integration of careers into each curriculum lesson, as well as the culminating career lessons, helped students understand the many different careers related to genetic testing and genetic research. Interviews and photos were provided with each career to help students connect that career with a real-world STEM professional.

Diversity of STEM individuals featured was a key consideration in lesson development, with an emphasis on women and individuals from backgrounds underrepresented in STEM, including people who were the first in their families to attend college. Many students do not have access to STEM professionals as role models. On seeing the photo of the veterinarian featured in lesson 2 of the introductory curriculum holding her infant son, one female student remarked to her teacher, “I didn’t know that you could be a veterinarian and have a family.” Providing diverse role models to all students can promote equity in the STEM fields. According to one Bio-ITEST teacher:

“I think being able to learn about various careers and how the learning each day is relevant to a particular career has been very valuable for my students. I struggle to incorporate this piece into my teaching on a regular basis. To have the career tie-in treated with intentionality and structured in a way that encourages students to really pay attention, was fabulous. They also were able to see an application of technology in science, and for my students, I think that was greatly valuable as well.”

Including stories or examples of STEM professionals throughout curriculum units may be an effective approach in a number of different STEM fields to promote career awareness among students.

Encourage Students to Use Authentic Scientific Tools and Approaches. The sense of self-efficacy that can arise from the ability to use the same tools used by practicing scientists may be a key factor in encouraging young people to consider a science career. In fact, the largest pre/postunit gain on a survey item was an indicator of self-efficacy (measuring understanding of how databases that store biological information are used in research settings). A Bio-ITEST teacher noted that, while the incorporation of career profiles and work responsibilities may have increased student career awareness, actually performing the types of computer analyses used by scientists is also compelling for students:

“I think the Bio-ITEST curriculum does a good job of generating career interest, but I suspect that simply doing the work of scientists is a compelling incentive for students to pursue this pathway.”

Emphasis on the practices of science via bioinformatics databases and tools is particularly timely, given the release of
A Framework for K–12 Science Education (NRC, 2011). The fact that the tools used (including BLAST, Cn3D, and FinchTV) are authentic bioinformatics tools used by scientists, not simplified or scaled-down versions made specifically for students, was also compelling:

“What they liked was that it (BLAST) wasn’t a ‘made for students program.’ They got the idea that they were able to use some of these tools that real researchers are using, that they can just look up stuff and find things. They really liked it, they asked, ‘Is this a thing that scientists use?’ They liked that idea that they were learning something real, not a made-up situation.”

While postunit measures of self-efficacy were significantly higher among students who experienced both the introductory and advanced units, only the advanced postunit survey showed significant gains in student engagement. This could be a result of a number of factors. Only the advanced unit contained hands-on wet-lab activities that, coupled with the computer-based activities, may have promoted greater engagement among these students. Students participating in the introductory unit were much more likely than advanced unit students to experience significant technology problems, such as limited access to computers and required programs (e.g., Cn3D was needed to view three-dimensional molecular images; described below in Anticipate Technology Challenges in the Classroom and Develop Potential Alternatives), which may have reduced measures of engagement. In addition, students who participated in the advanced unit spent more time on the Bio-ITEST activities (a total of 10–15 lessons for the introductory plus advanced unit students vs. four to eight lessons for the introductory unit only). It is important to note, however, that self-efficacy and engagement are intrinsically related to one another. When students are engaged with lessons or activities, they may be more motivated to improve their skills, leading to increases in self-efficacy. Conversely, a sense of self-efficacy can encourage students to explore subject matter more deeply, promoting greater engagement. Among students who participated in the introductory unit, the correlation between engagement and self-efficacy increased from $r = 0.50$ to $r = 0.63$ from preunit to postunit. Pre/postunit gains on survey items measuring engagement and self-efficacy were similarly correlated. In other words, students started the introductory unit having a positive relationship between self-efficacy and engagement (i.e., if they were already self-efficacious, then they were likely to also already be more engaged). Further, this relationship grew stronger over the course of the introductory unit intervention. If a student made gains in self-efficacy, he or she also tended to make gains in engagement. Importantly, this correlation works in either direction: if a student made gains in engagement, he or she also tended to make gains in self-efficacy. Because this is a correlational study, we cannot untangle the causal direction of the two variables and suspect that for some students self-efficacy may lead to greater engagement, while for other students engagement may lead to greater self-efficacy. Additional research is needed to disentangle these effects.

Whether by means of self-efficacy or engagement, or both, providing ample opportunities for students to use the authentic tools of a particular STEM field helps them understand what a career in that STEM field might entail. Using the “real” tools of the STEM field in lieu of “student versions” may increase both student interest in science content and a sense of self-efficacy.

Use Socio-Scientific Issues and Ethical Analysis to Increase Student Interest in Science Content. The National Education Science Standards (NRC, 1996) and A Framework for K–12 Science Education (NRC, 2011) emphasize not only science content and processes, but also the social contexts of science and the real-world decision making that students will face in their everyday lives. NWABR has successfully utilized socio-scientific issues and ethical discussion strategies to promote student interest in science content (Chowning, 2005, 2009a, 2009b). We have also shown that discussion of socio-scientific issues in science classes promotes higher-order justification skills among students and increases self-reported interest in and engagement with science content (Chowning et al., 2012). In our prior program evaluations, NWABR teachers have also reported increases in student interest in science-related material when it is framed within a socio-scientific narrative. Our prior work and the input of our curriculum-development teachers provided the foundation for our selection of genetic testing as the focus of the introductory unit and informed the inclusion of bioethics lessons in both curriculum units.

Among the two pre/postunit survey items measuring changes in student perceptions of relevance, neither of the changes was significant. However, students began the unit with fairly high scores measuring their interest in how science knowledge guides ethical decision making (a measure of relevance), and these scores changed little postunit. In addition, students reported significant gains on a related retrospective question evaluating their understanding of ethical issues in genetic testing and similar gains on a retrospective question about their understanding of the connection between biology content and issues they might face in their personal lives. The second relevance question, “I think it is important for people in our society to learn about science,” also had fairly high preunit scores. These scores changed little postunit, suggesting that students did in fact believe that having society learn about science is important, and exposure to the curriculum unit did not change those beliefs. In general, when students can see the connection between the science they learn in the classroom and real-world problems, engagement increases, and the material they are learning seems relevant (Shernoff et al., 2003; Siegel and Ranney, 2003). This may be particularly important for girls (Tucker et al., 2008; Modi et al., 2012), who have traditionally been underrepresented in STEM. During our research, we noted the key role of teachers in facilitating STEM career awareness; this role and its implications are discussed in detail elsewhere (Cohen et al., 2013).

Challenges for Program Implementation

Anticipate Technology Challenges in the Classroom and Develop Potential Alternatives. Some teachers were unable to implement particular lessons due (at least in part) to technology challenges, primarily lack of computer access for students and difficulty receiving school or district approval to install Cn3D and/or FinchTV on school computers. Many schools or districts require notification of the IT department weeks or months in advance of unit instruction to ensure computer availability and program installation. For some teachers, even
Balance the Challenges of Curriculum Fidelity versus Flexibility. While the evaluation did not document permutations in implementation, it is important to note that even for those who trialed most or all of the lessons, fidelity of implementation varied. Some teachers were constrained by school or district mandates on curriculum content, and thus did not teach courses in which the material would be relevant during that academic year, or believed that the material was too difficult for their introductory biology students. Some lessons were skipped, taught out of sequence, or modified due to the technical issues described above. For those who did implement the lessons, some noted that in some cases they skipped lessons entirely, chose to introduce only parts of a lesson, and/or replaced activities with their own innovations. In some instances, the results led to new ideas and approaches to lesson implementation, extension exercises, and other lesson revisions. For example, the mock job interview and social networking activities in the culminating career lessons were initially developed by field-test teachers as lesson permutations prior to their incorporation into the final version of the curriculum units.

Nonetheless, the gains in student career awareness, relevance, self-efficacy, and (post–advanced unit) engagement were encouraging, given the length of the intervention (usually six to nine lessons taught in a 1- to 2-wk period) and the challenges to implementation noted above.

One can view fidelity and flexibility as two sides of the same implementation coin. It is impossible to design a curriculum for implementation in varied educational settings that is “one size fits all,” and indeed, teacher innovation was important in refining the curriculum over time. However, flexibility and teacher adaptation of lesson components has the potential to dilute what may already be a short-term intervention, which makes measuring the outcomes of the intervention more challenging.

Consider Challenges in Program Evaluation and Areas for Additional Study. The survey instruments and career constructs utilized in this study provided a valuable means to assess program impacts on teachers and students; however, they have their limitations. The development of STEM identities among students is not a linear process, nor do all students experience to the same degree every component of the career development model that guided this study. While awareness of career possibilities is a necessary prerequisite for future pursuit of a given career, some students may be motivated to pursue additional STEM studies based on a sense of self-efficacy or mastery of the bioinformatics tools featured in the lessons (“I’m good at this, therefore I will try more of it”). Others, particularly girls, may be compelled by feelings of relevance (“this is important, this can help others”; Modi et al., 2012). The limited number of teachers who participated in the research study may have impacted the ability to detect a correlation between teacher change and student change. In addition, variations in fidelity of implementation could also influence these results. Finally, there may be some areas in which a correlation might not be expected. For example, students may be more comfortable with some aspects of technology than teachers are, so some areas in which teachers needed professional development and showed gains on survey items may not be the same areas in which change among students would be expected. Future studies would benefit from a greater number of teacher participants, as well as additional refinement of survey instruments, such as inclusion of a measure of engagement on the retrospective postintroductory unit survey. It would also be valuable to determine the generalizability of these findings to other STEM fields that utilize curriculum development and teacher professional development to promote student interest in STEM careers. For example, we believe that the effects of socio-scientific discussions on student interest in course content and careers in other STEM fields warrant further investigation.

CONCLUSION

According to the National Science Board, the number of STEM workers from 1980 to 2000 increased at an average annual rate of 4.2%, while the rate of STEM degrees acquired grew at only 1.5% annually (National Science Board, 2008). PCAST found “that the problem is not just a lack of proficiency among American students; there is also a lack of interest in STEM fields among many students” (PCAST, 2010). High school science curricula that explore real-world
problems and utilize authentic science tools appear to be an effective way to interest students in STEM content and promote self-efficacy. Additionally, incorporating ethical theory and discussing the socio-scientific issues arising from emerging genetic technologies appears to help students understand the relevance of the science material. Intentionally integrating career components into the curriculum helps acquaint students with the types of STEM careers available, the type of work each STEM professional performs, and the training and education requirements. Like many emerging areas of STEM, bioinformatics tools are complex and their use is challenging to teach. In addition, their user interfaces change often. However, utilizing the same bioinformatics tools that scientists use, such as the NCBI databases, BLAST, and Cn3D, promotes student interest and provides access to the wealth of biological data accumulated by scientists around the world. As one student aptly put it, “I had no idea that the general public has access to all of these databases and information.” Another Bio-ITEST student noted:

“Careers in science look more desirable than they did before, as now they are better explained …They don’t seem as tedious or difficult with the added features from bioinformatics.”

Familiarity with these tools will serve these students well if they pursue careers in STEM fields. For students who do not pursue careers in STEM, understanding the applications and limitations of bioinformatics tools and emerging genetic technologies will assist them in making informed decisions about medical advances they read about in the popular press or in the ballot booth.

ACKNOWLEDGMENTS

We thank Leena Pranikay, Joan Griswold, MIT, Reitha Weeks, PhD, and Jennifer M. Fang, PhD, for their assistance in program development and implementation and their review of the manuscript components. We also thank Lee Weigt, PhD, and Amy Driskell, PhD, of the National Museum of Natural History, Smithsonian Institution, for their assistance in the development of the DNA barcoding laboratory protocols for use with teachers and students. Thanks also to the scientist and student curriculum reviewers: Brian Fox, PhD, a senior scientist of computational biology at Novo Nordisk; Mette Peters, PhD, then a postdoctoral fellow in the Department of Biomedical and Health Informatics at the University of Washington; and Carol Farris, PhD, then a senior scientist of computational biology at Novo Nordisk; Mette Peters, PhD, then a director at the Institute of Translational Health Sciences at the University of Washington; and Carol Farris, PhD, then a postdoctoral fellow in the Department of Biomedical and Health Informatics at the University of Washington. Finally, we also thank the Bio-ITEST Advisory Board for their guidance and feedback and the National Museum of Natural History, Smithsonian Institution, for their assistance in the development of the DNA barcoding laboratory protocols for use with teachers and students. Thanks also to the scientist and student curriculum reviewers: Brian Fox, PhD, a senior scientist of computational biology at Novo Nordisk; Mette Peters, PhD, then a postdoctoral fellow in the Department of Biomedical and Health Informatics at the University of Washington; and Carol Farris, PhD, then a senior scientist of computational biology at Novo Nordisk; Mette Peters, PhD, then a director at the Institute of Translational Health Sciences at the University of Washington; and Carol Farris, PhD, then a postdoctoral fellow in the Department of Biomedical and Health Informatics at the University of Washington. Finally, we also thank the Bio-ITEST Advisory Board for their guidance and feedback and the teachers and students who made this research possible. This research was supported by the NSF ITEST program (DRL 0833779). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES


Chowning JT (2005). How to have a successful science and ethics discussion. Sci Teach 72, 46.


President’s Council of Advisors on Science and Technology (2010). Prepare and Inspire: K–12 Education in Science, Technology, Engineering and Math (STEM) for America’s Future, Washington, DC, vi.


We report on the development of a life sciences curriculum, targeted to undergraduate students, which was modeled after a commercially available physics curriculum and based on aspects of how people learn. Our paper describes the collaborative development process and necessary modifications required to apply a physics pedagogical model in a life sciences context. While some approaches were easily adapted, others provided significant challenges. Among these challenges were: representations of energy, introducing definitions, the placement of Scientists’ Ideas, and the replicability of data. In modifying the curriculum to address these challenges, we have come to see them as speaking to deeper differences between the disciplines, namely that introductory physics—for example, Newton’s laws, magnetism, light—is a science of pairwise interaction, while introductory biology—for example, photosynthesis, evolution, cycling of matter in ecosystems—is a science of linked processes, and we suggest that this is how the two disciplines are presented in introductory classes. We illustrate this tension through an analysis of our adaptations of the physics curriculum for instruction on the cycling of matter and energy; we show that modifications of the physics curriculum to address the biological framework promotes strong gains in student understanding of these topics, as evidenced by analysis of student work.

INTRODUCTION

In 1910, the educator and philosopher John Dewey noted that “science teaching has suffered because science has been so frequently presented just as so much ready-made knowledge, so much subject-matter of fact and law, rather than as the effective method of inquiry into any subject-matter” (cited in Archambault, 1964, p. 182). Nearly 100 yr later, this sentiment that science is best learned through active inquiry and exploration rather than a traditional lecture approach is echoed by the reports and recommendations of the American Association for the Advancement of Science (AAAS, 1997, 2011), the National Science Foundation (NSF, 1996), and the National Research Council (NRC; Bransford et al., 1999). Unfortunately, the available textbooks and curricula commonly stress memorization over the deep conceptual understanding that develops when students engage in inquiry-based activities and discussions that help them construct their own understanding of science concepts.
Reform in these areas is particularly critical when considering the training of our nation’s future teachers. Teachers model their own teaching after the classroom experiences they had as learners more than on the theory or even the classroom experiences they encounter in teacher education programs (Grossman, 1991). Unfortunately, the contrast between what is expected of future teachers in their K–12 classrooms and what they experience in content and instruction in typical college or university science courses can be quite striking (Darling-Hammond and Bransford, 2005). Thus, a targeted approach to teacher education must begin in their undergraduate science preparation, and prospective teachers should be taught science in a manner that replicates the inquiry strategies and active learning that we hope they will employ in their own classrooms (McDermott, 2006). Many elementary teachers, even experienced ones, are uncomfortable teaching science for a variety of reasons, including poor scientific literacy, negative attitudes toward science, and the belief that science is difficult to teach (see van Aalderen-Smeets et al., 2012 and references therein). Because elementary teachers feel the least prepared to teach science compared with other subjects (Fulp, 2002; Dorph et al., 2007), it is crucial that any undergraduate science classes they do experience employ the inquiry strategies and constructivist approaches that characterize best practices in science education. Indeed, modeling effective inquiry strategies can change the practice of elementary teachers, causing them to use more inquiry in their own classrooms (Staples, 2002).

Preparing preservice elementary teachers to teach science effectively was one important aspect of the North Cascades and Olympic Science Partnership (NCOSP), an NSF-funded Math–Science Partnership housed at Western Washington University (WWU). One way we addressed this was by designing a year-long sequence of three science courses (in physical science, life sciences, and earth science) targeted to elementary education students completing their credentials at WWU but open to all undergraduates. Initially, we sought published science curricula that adopted a student-centered, constructivist, active, and inquiry-based approach to deepen science content knowledge in a manner appropriate for elementary education undergraduates.

More than other science disciplines, the physics education community has developed curricula that meet these criteria, including Physics by Inquiry (McDermott, 1996), Modeling Instruction (Hestenes, 1987; http://modelinginstruction.org), Investigative Science Learning Environments (Etkina and Van Heuvelen, 2007; Etkina et al., 2010), and Physics for Elementary Teachers (now called Physics and Everyday Thinking [PET], Goldberg et al., 2005). Although all of these curricula were developed based on research on how people learn, they differ in their pedagogical approaches and in their target audiences. Of the curricula available at the time, PET best suited our needs for the physics course of our three-course sequence, because its pedagogy fit our class structure (small, laboratory-based classes) and its target audience is elementary teachers. However, as we began preparing for the other science courses in the sequence (those focused on life and earth sciences), we were faced with a lack of cohesive, constructivist curricula appropriate for our needs. In life sciences, although there are several examples of reformed introductory courses, these tend to focus on implementing innovative instructional strategies into large classes (e.g., Frederichsen, 2001; Lawson et al., 2002; Chaplin and Manske, 2005; Hoskins et al., 2007; Uekert et al., 2011). We required materials for a smaller, lab-based course that modeled pedagogy that could be used by preservice elementary teachers. Thus, to provide a cohesive experience for students taking all three classes, we decided to develop a life sciences curriculum using PET as a model. (An earth science class using PET as a model was also developed by other faculty associated with NCOSP.)

As a discipline, biology presents some unique challenges to student learning. Recent work in the field of developmental cognitive psychology has elucidated why some common misconceptions persist, given how the human mind functions. Coley and Tanner (2012) sum these up as “cognitive construals,” which they define as “informal, intuitive ways of thinking about the world.” Cognitive construals are ways in which people process, understand, and make decisions about information, and some cognitive construals have the capability to influence how a student learns biology. The authors argue that three are particularly important to biology education. The first is teleological thinking, a type of causal reasoning in which students assign a goal or purpose to a process. The common misconception that plants produce oxygen because animals need it is an example of teleological thinking. This type of thinking is common and useful in everyday thought, because it allows us to interpret events and behaviors; it is therefore persistent even when contradictory evidence is available (Kelemen, 1999). The second cognitive construal identified by Coley and Tanner (2012) as particularly troublesome for learning biology is essentialist thinking, in which students assign an essential property, or “essence,” to members of a group or category along with other summary information about that group. For instance, many students consider DNA to be the essential property that makes a cell a cell. Such thinking can lead to misconceptions, such as thinking that cells in a body must have different DNA, because the cells look different from one another. Essentialist thinking can also lead a learner to believe that there is little or no variation in a group. Finally, the third cognitive construal that is important to biology education is anthropocentric thinking. In this case, students use analogies about humans to try to make sense of unfamiliar biological phenomena or processes. The common misconception that plants take in food molecules through their roots is an example of this type of thinking.

Students also have difficulty understanding concepts at the different scales (atomic, cellular, organismal, ecosystem) that apply to biological phenomena. Hartley et al. (2011) argue that this is partly due to the pervasiveness of informal reasoning (reasoning that uses simple “actors” that cause events to happen in the presence of “enablers”) as opposed to principle-based reasoning, which relies on fundamental laws to explain phenomena. To fully understand many concepts in biology, students must be able to use principle-based reasoning to move across scales. For example, to understand how energy and matter move through living systems, students must be able to use the laws of conservation of energy and matter to move from the atomic to the ecosystem level (Wilson et al., 2006; Hartley et al., 2011).

Overcoming these challenges will require a change in how biology is taught to undergraduates, and there have been several recent reports calling for such a change (for an overview, see Labov et al., 2010). In particular, Vision and Change in Undergraduate Biology: A Call to Action (AAAS, 2011) outlines...
best practices in teaching biology, practices that are still rare in most undergraduate biology classrooms, in which lecture is the norm, and content is covered at a rapid rate. The authors of Vision and Change advocate covering fewer concepts in greater depth, making learning goals for core concepts explicit to students, and integrating science process skills throughout the curriculum. They also advocate “student-centered learning,” which involves students as active learners in all classes, uses different types of instruction (including lecture), and integrates multiple forms of ongoing assessment throughout a course. These changes are necessary, especially given the rapid advances in biology research and the increasing intersection of biology with other disciplines.

In this paper, we describe the process we used to develop a reformed life sciences curriculum that uses a physics curriculum model, addresses the challenges inherent to learning biology, and responds to the call to change how undergraduate biology is taught. We briefly describe the resulting materials and then discuss the challenges we faced in adapting a life sciences curriculum based on a physics model. Finally, we illustrate the effectiveness of our curriculum, using a well-established, open-ended assessment that was given to students in several different courses, including some not using our curriculum.

CURRICULUM DESIGN AND DEVELOPMENT

The life sciences curriculum, called Life Science and Everyday Thinking (LSET), was initially developed, beginning in September 2004, by a group of higher-education faculty from a regional university (WWU), three local community colleges that provide the majority of transfer students to WWU (Evett Community College, Skagit Valley College, and Whatcom Community College), and the Northwest Indian College. The initial curriculum was intended for a 10-wk course and was subsequently revised, beginning in February 2010, to a 16-wk curriculum for institutions on the semester system. This last revision was done in collaboration with colleagues at California State University, Chico (CSU, Chico) and with middle- and high-school teachers from the Bellingham, Washington, area.

Our target audience was preservice elementary teachers, although we wanted the curriculum to be appropriate for all nonscience majors, as many of the students taking the courses at the community colleges did not intend to become teachers. WWU has a large teacher preparation program and, in the past, our preservice elementary teachers took a minimal number of “introductory” science courses from a broad menu of topics and disciplines. These courses were often survey courses that provided a superficial treatment of many topics within a discipline, and lecturing was the predominant instructional strategy used. Thus, our elementary education students were not being provided the opportunity to develop deep, conceptual understanding of relevant content, nor were they able to observe instruction that applied the recommendations of research findings on how people learn. Our students now take a three-course sequence that includes PET, LSET, and the earth science curriculum also developed through NCOSP. (A fourth course on chemistry has just been developed.) CSU, Chico also has a large population of future teachers who previously had little coherent science preparation targeted to their needs—the development of the semester-long version of LSET was a response to meet those needs.

To link the courses in the sequence, LSET explicitly uses the same pedagogy as PET. We also linked the courses thematically, by using the flow of matter and energy in living systems as focal points for LSET. Overall, LSET covers the one-way flow of energy from sunlight captured by plants through the trophic levels, with heat loss throughout. These concepts are contrasted with the cyclical flow of matter, focusing on carbon. Over five chapters, the curriculum presents these concepts at the cellular, organismal, and ecosystem levels. It also has two chapters (one on genetics and one on evolution) less directly related to this main theme that we considered necessary to include in the curriculum due to elementary teacher preparation guidelines in several states. The full curriculum can be covered in a semester. On the quarter system, instructors have the choice of two coherent paths: one focusing on the organismal/ecosystem concepts and one focusing on the organismal/cellular concepts. Figure 1 illustrates the range of concepts covered by LSET.

A “backward design” (Wiggins and McTighe, 1998) was used to develop the curriculum. Thus, we first identified what we wanted our students to know or be able to do at the end of a chapter and developed assessment items around these learning goals. Then we developed the chapter to better enable students to construct understanding toward those goals.

Figure 2 provides a conceptual model of the specific steps we used during the development process. The process began with identifying the key concepts (we refer to them as “big ideas”) that students need in order to understand the topics covered in the course. Content experts from the university and community colleges consulted the state and national science standards in circulation at the time of curriculum development, including the Grade Level Expectations for Washington State Essential Academic Learning Requirements (Washington State Office of the Superintendent of Public Instruction [WOSPI], 2005), the National Science Education Standards (NRC, 1996), and the Benchmarks for Science Literacy and Atlas of Science Literacy (AAAS, 1993, 2000) for identification of these big ideas. They also relied on their own training. For example, two of the big ideas found in chapter 4 of LSET (“How do matter and energy cycle in living systems?”) are:

- Energy flows through an ecosystem, entering mostly as light, passing through as chemical energy in organic compounds, and exiting as heat; and
- Matter, in the form of essential chemical elements, is recycled within an ecosystem by decomposers, which decompose organic material and return elements to reservoirs.

These big ideas are echoed in the National Science Education Standards (NRC, 1996), which state the by the end of high school students should understand that:

As matter and energy flows through different levels of organization of living systems—cells, organs, organisms, communities—and between living systems and the physical environment, chemical elements are recombined in different ways. Each recombination results in storage and dissipation of energy into the environment as heat. Matter and energy are conserved in each change. (pp. 186–187)
The Benchmarks also incorporate this idea, but over a range of age groups. By middle school, students should understand that:

- Over a long time, matter is transferred from one organism to another repeatedly and between organisms and their physical environment. As in all material systems, the total amount of matter remains constant, even though its form and location change.
- Energy can change from one form to another in living things. (p. 120)

and by the end of high school they should understand how these concepts apply to populations and ecosystems:

- At times, environmental conditions are such that land and marine organisms reproduce and grow faster than they die and decompose to simple carbon containing molecules that are returned to the environment. Over time, layers of energy-rich organic material inside the earth have been chemically changed into great coal beds and oil pools.
- The chemical elements that make up the molecules of living things pass through food webs and are combined and recombined in different ways. At each link in a food web, some energy is stored in newly made structures but much is dissipated into the environment. Continual input of energy from sunlight keeps the process going. (p. 121)

This big idea is also found in state standards documents. The Washington State Essential Academic Learning Requirements (WSOSPI, 2005) state that by the end of high school students should understand that

*Matter cycles and energy flows through living and non-living components in ecosystems. The transfer of matter and energy is important for maintaining the health and sustainability of an ecosystem.* (p. 98)

Finally, this big idea is found in one of the core concepts for biological literacy outlined in the recently released *Vision and Change in Undergraduate Biology: A Call to Action* (AAAS, 2011):

Biological systems grow and change by processes based upon chemical transformation pathways and are governed by the laws of thermodynamics. (p. 13)

This tight relationship between the LSET big ideas and the state and national science standards is found throughout the LSET curriculum.

Once the big ideas were established (a process that took several months), we used them to identify the content critical to understanding them. With this in mind, we outlined the smaller ideas, or “subideas,” necessary to construct understanding of the big ideas and how those ideas link together. Another set of ideas was developed that address the skills and practices inherent to the practice of science regardless of discipline, for example, experimental design, data presentation, and identification of patterns. Our goal was that no individual activities should address these “process” ideas on their own; rather, these skills and practices should be embedded within developing biology content, as part of what it means to do and learn science.

Once the big ideas were identified, the curriculum was divided into chapters and individual or pairs of faculty members agreed to work on a chapter. Summative assessments were written. Common initial student ideas were identified for each of the big ideas (AAAS, 1993; Driver *et al.*, 1994), and the curriculum was developed to specifically expose and confront these preconceptions.

For the curriculum, we relied on commonly used experiments, but we linked them together in a manner that reflects how students best learn science. Rather than having students complete experiments in a “cookbook” manner, each chapter was subdivided into activities, and each activity was developed to follow a learning cycle. For example, students are...
led to “discover” that plants undergo both photosynthesis and cellular respiration by first recording their initial ideas about what will happen to oxygen and carbon dioxide levels in closed chambers containing plants when the chambers are in the light and in the dark. They discuss their initial ideas with one another and then present their ideas to the class using a whiteboard. They then measure oxygen and carbon dioxide levels and record their data in their notebooks and on a class graph. Throughout the experiment, they answer questions that deliberately link their results to earlier measurements of gas levels in chambers containing animals (the concept of cellular respiration was initially developed in a previous chapter). In the next experiment, they observe infrared pictures of a corpse flower to conclude that plants give off heat, again answering questions that link their observations to the products of cellular respiration. At the end of the activity, students revisit their initial ideas and document how their thinking has changed. Thus, rather than just following a procedure and recording data, students’ preconceptions are elicited, conceptual understanding is constructed, and students engage in metacognition at the end of the activity. This learning cycle is repeated for each activity within a chapter.

The structure of a chapter is illustrated in Figure 3, which uses chapter 3 (“What Is Food for Plants?”) as an example of how the different components within a chapter are organized. Briefly, at the beginning of each chapter, the purpose is stated, and the students are prompted to complete a formative assessment that addresses the big ideas of the chapter. They discuss their ideas in small groups and then share those ideas with the class using whiteboards. This allows instructors and all class participants to hear the variety of initial ideas. Students then complete a series of activities designed to specifically address common ideas and to allow students to construct knowledge in a sequential manner, as described above. In general, the experiments include laboratory activities, thought experiments, and exercises using paper and computer models. Throughout these experiments, students are required to make predictions, gather data, and draw conclusions. At the end of each activity, students are prompted to explicitly reconsider ideas held before the activity and to document any change in their thinking. At times when “telling” is required, material is presented as Scientists’ Ideas, in which students read factual information about the concepts they have been investigating and link their observations and results to that information. Homework is assigned after most activities, and it is done individually to allow instructors to check for individual student understanding of the key concepts from the activity. At the end of each chapter, students revisit the ideas they held in the formative assessment at the beginning of the chapter and consider how their thinking has changed. They also participate in a group discussion about the big ideas of the chapter, and this is a chance for the instructor to guide students to link concepts into a broader picture. Finally, each chapter contains a Learning about Learning activity, in which students consider different aspects of how people learn as they apply to the LSET curriculum and to elementary classrooms.

Comparing the Physics and Biology Curricula

Table 1 compares the learning principles used in PET (Goldberg et al., 2005) with the learning principles in our curriculum. These principles are derived from research in cognitive science and science education and are not unique to physics; further details on their origins are described in Goldberg et al. (2010). It was relatively straightforward to incorporate these elements into the biology curriculum, as noted in the rightmost column of Table 1. There were occasional challenges in developing a good elicitation question, finding appropriate tools, and scaffolding skills, but these challenges did not seem unique to biology, suggesting (as expected) that these learning principles are relatively independent of scientific content. Nonetheless, there were elements of the PET curriculum that required a significant rethinking. In particular, challenges included:

- Using representations of energy developed for physics contexts (e.g., a diagram showing how a block gains and loses energy as it is pushed) in biological contexts (e.g., a diagram showing how a tree gains and loses energy);
Figure 3. Overview of the LSET curriculum using chapter 3 as an example of the curriculum structure. Each chapter and each activity within the chapter are organized with the same learning cycle components.

- Defining fundamental concepts (life, species, evolution);
- Handling anomalous results; and,
- Incorporating scientists’ ideas at the end of a chapter.

We have come to characterize these challenges as reflecting a difference between the core ideas presented in introductory physics and biology. In particular, we interpret these challenges through the following lens: **introductory physics is often presented as a science of pairwise interactions; introductory biology is often taught as a science of linked processes.**

In introductory physics, pairwise interactions (e.g., the free body diagram and the PET energy transfer diagram) underlie explanations and predictions for phenomena. In introductory biology, understanding the types of core processes, including their inputs, outputs, and functions, is the building block upon which explanations of biological phenomena rest. And, we claim, this difference accounts for the primary challenges we faced in developing a biology curriculum, LSET, based on a pedagogy successful in a physics context, PET.
In the following sections, we offer support for these claims by examining the Framework for K–12 Science Education (NRC, 2012), LSET, and PET, using cellular respiration as a brief case study. It is worth noting that these orientations (pairwise vs. linked processes) are not necessarily intrinsic to the disciplines: expertise in biology often requires an understanding of the interactions underlying processes. Similarly, expertise in physics requires being able to move fluidly from simple pairwise interactions to aggregate phenomena.

### Cellular Respiration as a Biological Process

Take, for example, a core topic in introductory biology: aerobic cellular respiration. In our view, the key biological idea of aerobic cellular respiration is that it is the process by which chemical energy, stored in the arrangement of food molecules (such as glucose), is transferred to ATP, allowing a living organism to use the energy in ATP to carry out the necessary processes of life. For this transfer to take place, the atoms of molecular oxygen and glucose are rearranged into water and carbon dioxide, which are then typically released from the cell. This process is one that connects producers, consumers, and their abiotic environment. It helps to explain such key observations and questions in biology as: Why do we need to breathe? What is the purpose of food? What kinds of substances give us energy? What kinds of organisms need oxygen? What role does breathing play in losing weight? How are organisms connected to one another and to their environment?

This description of the core ideas of cellular respiration is not an idiosyncrasy of the LSET curriculum, but is echoed in the new Framework for K–12 Science Education (NRC, 2012), which notes that (by grade 12) students should understand the following ideas related to matter and energy flows as part of the disciplinary core ideas for the life sciences:

> Aerobic cellular respiration is a chemical process in which the bonds of food molecules and oxygen molecules are broken and new compounds are formed that can transport energy to muscles … Cellular respiration also releases the energy needed to maintain body temperature despite ongoing energy loss to the surrounding environment. (p. 148)

In both LSET and the Framework, the core idea of cellular respiration is a process, rather than an interaction, and our attention is drawn to characterizations of matter and energy at the beginnings and endpoints of processes, not the beginnings and endpoints of pairwise interactions. At the

---

**Table 1.** A comparison of learning principles used in PET and LSETa

<table>
<thead>
<tr>
<th>Learning principles</th>
<th>How PET applies principles</th>
<th>How LSET applies principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students’ prior knowledge influences their learning.</td>
<td>Student’s initial ideas are elicited at the beginning of most activities. Activities make use of students’ intuitive ideas and build on previously constructed ideas.</td>
<td>As with PET, most activities begin with initial ideas. For example, students have an intuitive sense that there are no cells the size of a watermelon, that animals lose weight by excreting it, and that plants must take in solid materials to grow.</td>
</tr>
<tr>
<td>Students’ knowledge may be resistant and is often at odds with science ideas.</td>
<td>Activities are explicitly designed to elicit and then address commonly held ideas. Students are asked to revisit their initial ideas at the end of activities.</td>
<td>The ideas in LSET, as in PET, are explicitly elicited and engaged through the curriculum and revisited at the end of each chapter.</td>
</tr>
<tr>
<td>Students construct knowledge gradually in a complex process requiring multiple exposures.</td>
<td>Activities within and across chapters build on one another. Particularly resistant ideas are addressed explicitly several times in different contexts.</td>
<td>Not only does LSET build through the entire semester—tackling the theme of “what is life” and flows of matter and energy in living systems—but it builds on ideas developed in PET.</td>
</tr>
<tr>
<td>Complex skills can be scaffolded and modeled over time.</td>
<td>The skills of constructing and evaluating a scientific explanation are first introduced with a lot of support and structure. This support fades over the length of the course.</td>
<td>These skills—first introduced in PET—are continued in LSET, developing the ideas over two semesters (or quarters). In addition, exploring the flow of matter and energy is first introduced in relatively simple systems and builds to consider an entire ecosystem through this curriculum.</td>
</tr>
<tr>
<td>Students’ learning is mediated by social interactions.</td>
<td>Students engage in cooperative learning by working through the activities in small groups. The end of every activity also includes class discussion of some initial ideas and summarizing questions.</td>
<td>In LSET, as in PET, students work closely on investigations in small groups and then use whole-class discussions to discuss and evaluate the claims they can now make.</td>
</tr>
<tr>
<td>Interaction with tools is critical to learning.</td>
<td>Whenever possible, students perform hands-on experiments to gather evidence. Computer simulations, video, and instructor demonstrations extend this experience.</td>
<td>Tools, including CO2 detectors, thermometers, microscopes, whiteboards, online data, fossils, and those used in student-generated experiments, are central to the curriculum.</td>
</tr>
</tbody>
</table>

aColumns 1 and 2 from Goldberg et al. (2010).
start of cellular respiration, there are certain forms of energy present in the arrangement of molecules; at the end of this process, the energy is located in different molecules. Other questions may arise: In what way do food and oxygen molecules “break,” form new molecules, and maintain temperature? Why should these interactions form new molecules that can “transport energy”? How does a molecule “transport energy”? Such questions are legitimate scientific questions—however, they are not presented as biological questions. Instead, they point toward the (usually pairwise) interactions that underlie biological processes and are not part of what is considered introductory biology content.

By way of comparison, consider the physical science disciplinary core ideas presented in the Framework that address cellular respiration (among other biological processes):

- A variety of multistage physical and chemical processes in living organisms, particularly within their cells, account for the transport and transfer (release or uptake) of energy needed for life functions. (p. 130)

Such a statement is so vague as to be virtually useless as an explanation; instead, we view this description as an epistemological claim: students should understand that underlying the process of cellular respiration that biology claims as a core idea are a multitude of physical interactions that account for those biological phenomena. This core idea, then, regards the kind of explanations offered in the different disciplines: physical science is nested at a more fundamental level, attending to multistage interactions.

How, then, would physics address the questions of transfers of energy in cellular respiration?

Transfers of Energy as Pairwise Interactions

The PET curriculum addresses how to account for the changes in energy that take place during chemical reactions as the attractive and repulsive components of atoms rearrange during processes such as cellular respiration by guiding students to construct the following idea:

- If the mutual interactions between the components of a system are attractive, then as the average distance between the components increases, the potential energy of the system increases also. If the mutual interactions between the components of a system are repulsive, then as the average distance between the components increases, the potential energy of the system decreases. (Goldberg et al., 2005, p. 3-89 in Scientists’ Ideas)

And, again, this idea is not an idiosyncrasy of the curriculum, but is an idea that is echoed in the Framework. The physical science core disciplinary ideas that address how arrangements of atoms might store energy claims that:

- When two objects interacting through a force field change relative position, the energy stored in the force field is changed. Each force between the two interacting objects acts in the direction such that motion in that direction would reduce the energy in the force field between the objects. (NRC, 2012, p. 127)

The two statements convey the same idea: One interaction between two objects leads to a predictable change in energy. These descriptions provide a mechanism by which two interacting objects gain or lose energy—addressing why a ball thrown upward slows down as it rises, or why attracting magnets accelerate as they come together. Using a PET approach in the LSET curriculum, then, we would describe both the food–oxygen pairing as the container of chemical potential energy and the attractive forces that convert chemical potential to kinetic energy as the food and oxygen are brought closer together.

It is reasonable to assume that the topics in introductory physics are also explained by more fundamental interactions than are presented in the typical introductory course—that is, much like introductory biology does not concern itself with the detailed, molecular mechanisms of the Krebs cycle to establish core ideas regarding the flow of matter and energy, there may be a multitude of mechanisms underlying the means by which two interacting objects change their energy that are not the part of the canon of introductory physics. However, unlike the Framework’s statement acknowledging that physical processes underlie biological processes, there is no analogous statement in physics; the Framework does not claim, for example, that

- A variety of multistage subatomic and quantum processes account for the transfer of energy between two interacting objects.

Though this is in fact true, introductory physics—at least within the Framework and PET—presents itself as being a base-level, fundamental mechanism for the sciences, dealing with interactions between two irreducible objects. (It is not that the objects are fundamentally irreducible, but that they are presented as such.)

Effects on Curriculum Development

During the development of LSET, the difference between processes and interactions manifested itself in the following ways:

- Using representations to track the flow energy;
- Defining fundamental concepts;
- Handling anomalous results;
- Incorporating scientists’ ideas.

In the following sections, we detail how the difference between interactions and processes presented challenges for these topics, how we addressed them in our curriculum, and the results on student understanding of the flow of matter and energy in ecosystems.

Representations to Track the Flow of Energy

PET and LSET both frame much of their instruction on characterizing and following the transfer and transformation of energy through systems, whether they are living systems (cells, organisms and ecosystems) or physical systems (colliding carts, interacting magnets, or heated gas). Because of this, we had hoped to use the energy diagrams from PET to characterize energy flows in living systems.

For example, a hand pushing a cart at constant speed along a surface might be represented as shown in Figure 4. This energy diagram depicts all of the relevant interacting objects by linking the objects with an arrow; for each interaction (labeled by an underline), there is a transfer of energy, described by the form of energy in the arrow. This transfer affects the amount of energy present in each object, as indicated in the bubble below the object; energy decreases in the source and
increases in the receiver. (If two objects interact via more than one type of interaction, there would more than one arrow linking the objects.)

When seeking to describe the discrete interactions through which energy transfers and transforms during, say, cellular respiration, the sheer number of interactions is problematic. Moreover, in conversations between biologists and physicists, we found that such moment-to-moment, fine-grained descriptions of energy transfers and transformations were not part of the basic canon of introductory biology. A physicist, for example, would find it problematic to claim that energy “in” the bonds of glucose is transferred to the bonds of ATP and would locate that energy instead in the broader glucose–oxygen system, and yet, rather than claiming that biology gets it wrong, we have come to interpret this as attending to a level of mechanism that is not addressed by introductory biology (for a more detailed discussion on chemical bonds and energy in biological systems, see Redish [2012]; for an analysis of how biologists and physicists define and describe these concepts differently, see Hartley et al. [2012]). That is, we do not wish to teach biology as physics, and therefore must modify the representational format to better capture the kinds of questions and answers that biologists offer when considering the flow of energy.

So instead of calling students’ attentions to the steps of glycolysis/Krebs cycle—the pairwise interactions between molecules—we modified these energy diagrams to describe significant biological processes. The representations still follow the source/receiver path for energy, but the arrows represent particular kinds of life processes, as shown in Figure 5, rather than interactions.

Table 2 summarizes how the physics curriculum attends to particular pairwise interactions and how the biology curriculum attends to biological processes in tracking the flow of energy by describing a representative set of interactions (physics) and processes (biology).

**Defining Fundamental Concepts**

A second difference between the two curricula lies in the role that definitions play. Throughout the physics curriculum, core concepts—for example, kinetic energy, thermal energy, and mechanical energy—are defined for students. Within LSET, we chose to ask students to construct definitions for many key terms. For example, the term “food” is used without a strict definition when discussing animals, but when we begin to discuss what counts as food for a plant, students struggle with how to best define the term. Through explorations involving oxygen and carbon dioxide sensors, with plants in the light and plants in the dark, students consider how plants acquire the materials for growth and energy and use these ideas to construct their own definitions for food. For many classes, this results in a debate in which students argue whether or not carbon dioxide and sunlight should be considered food, or whether a material can only be considered food when the matter and energy are together in a single food molecule.

Within biology, determining what should be considered food is a question that ties to the role that food molecules play in ecosystems, organisms, and cells. It is only through understanding the role of certain molecules as building blocks and energy providers for all life processes that our definition of

![Figure 4](image1.png) Energy diagram from PET for a hand pushing a cart at constant speed. The diagram tracks energy transfers and transformations through a sequence of pairwise interactions (e.g., contact push–pull, heat conduction, and infrared interactions) between objects.

![Figure 5](image2.png) Energy diagram from LSET for food being used for energy. The diagram tracks energy through a series of linked processes (e.g., ingestion, cellular respiration, cell work) that transfer and transform energy.

Vol. 12, Highlights of 2013 121
food makes sense. Defining terms such as “food,” “life,” and “species” requires an understanding of the broader processes in which these terms have their meaning (cellular respiration, ecosystems, and evolution). That is, many key definitions in biology are so tied to the processes in which they derive their meaning that defining the terms outside of the context of the process (and before students have investigated and understood the process) is problematic. What makes a producer a producer is its ability to create food from nonfood molecules through photosynthesis; what makes food “food” is its role in growth and cellular respiration. Absent understandings of these key processes (photosynthesis, growth, and cellular respiration), such definitions are meaningless, so constructing definitions goes hand-in-hand with inquiry into biological processes.

By way of comparison, PET begins the curriculum with the following definition (the second activity of the first chapter; the first activity familiarizes students with motion detectors and their graphs):

Scientists associate a form of energy with the motion of an object—they call it kinetic energy. The faster an object is moving, the more kinetic energy it possesses. (Of course, an object at rest possesses no kinetic energy.) Thus, as the speed of an object changes, the amount of kinetic energy it has also changes. (Goldberg et al., 2005, p. 1-21, emphasis in the original)

With this definition in place, students investigate which kinds of interactions with which kinds of objects cause changes in kinetic energy. The interactions are so simple—pitting a single hand interacting with a cart on a track—that calling attention to one aspect of the interaction is not “giving away” a finding, nor does it require understanding the broader framework of energy to make sense of this definition. Thus, in PET, definitions are not student-constructed concepts, rather, they are instructor-provided vocabulary.

### Anomalous Results

Unanticipated laboratory results are common in school science labs; in introductory physics, however, troubleshooting is often straightforward. Our attention in physics is on constraining the system as much as possible to focus on a single, pairwise interaction. Because of this, inconsistent results are generally a result of student error and not due to variability inherent to all experimental work. Furthermore, experiments involving rolling carts, colliding objects, dropped balls, and attracting magnets can easily be replicated and consensus quickly reached. As a result, students in the PET course frequently pursue investigations in small groups with a chance to share results at the end of the lab and with infrequent problems in achieving anticipated results.

In biology, however, the multitude of interactions underlying the processes we investigate (photosynthesis, decomposition, growth, cellular respiration) are so complex, and involve live organisms responding with inherent variation, that unanticipated results are frequently due to variability that cannot be controlled in our lab setting and are not due to errors in implementing procedures. We quickly found that small-group work in teams of four led to inconsistent results and required groups to compile and discuss data as a whole class before proceeding. Instructors must actively facilitate these discussions to call attention to how variability may play a role, what kinds of trends we are seeing, whether or not those trends make sense, and whether we can account for the variation.

### Incorporating Scientists’ Ideas

In contrast to typical textbooks for introductory science, and consistent with best practices for science education (Bransford et al., 1999), both PET and LSET ask students to construct their own ideas through investigations and discussions with peers. To present canonical scientific descriptions of these ideas, students compare their ideas with scientists’ ideas. In PET, this happens at the end of each chapter, where students generally find that the ideas they have constructed are congruous with scientific principles. In LSET, however, we found that there were details that could not easily be constructed by students through laboratory investigations; certain background information needed to be presented earlier to enable students to move forward through the curriculum. For example, in chapter 2 ATP is introduced without a laboratory investigation, which might allow students to discover for themselves its role in biological systems, and it is also introduced before the end of the chapter. Students are asked...
to make sense of the idea and explore why this intermediate molecule is necessary to link cellular respiration and cell work. It was necessary to “tell” about this idea in the middle of the chapter, because the concept was necessary for building understanding of upcoming concepts, yet it was difficult to allow students to “discover” it for themselves.

Again, we can attribute this difference to the nature of the two courses; in a course attending to pairwise interactions, there are few hidden mechanisms to introduce. Within introductory biology, some facility with the mechanisms underlying the key processes is useful, but not always feasible or practical to introduce through student-centered investigations, especially when the mechanisms are molecular in nature and are thus difficult to investigate directly in an introductory course. In the case of ATP, we relied on an analogy (one of miniature rechargeable batteries) to introduce the concept through a Scientists’ Ideas section. Introducing those ideas early on in the chapter proved useful.

Assessing LSET

PET has proven itself as a curriculum that promotes strong understanding of core ideas surrounding energy in physics (Goldberg et al., 2010). A key question in evaluating the LSET curriculum centers around whether this curriculum—modeled on PET but with biological content and with changes in pedagogy necessitated by the shift to a processes view of science—would have similar success.

By way of example, the following assessment question exemplifies the big picture understanding that introductory biology courses seek:

Grandma Johnson had very sentimental feelings toward Johnson Canyon, Utah, where she and her late husband had honeymooned long ago. Because of these feelings, when she died she requested to be buried under a creosote bush in the canyon. Describe below the path of a carbon atom from Grandma Johnson’s remains, to inside the leg muscle of a coyote. NOTE: The coyote does not dig up and consume any part of Grandma Johnson’s remains. (Ebert-May et al., 2003, p. 1224)

This question requires students to integrate and apply their ideas about many linked biological processes—decomposition, photosynthesis, food webs, and carbon cycling. We know that students often have misconceptions about these topics. For instance, many students initially believe that decomposition is a natural physical change that dead matter undergoes, in which wind and water break organic matter down into smaller pieces, as though it went through a paper shredder. Our goals were for students to understand that decomposers actively and chemically change the dead material into new, small molecules through cellular respiration (Smith and Anderson, 1986). Similarly, many students believe that photosynthesis provides energy for plants to take in organic molecules such as glucose and amino acids through their roots, rather than serving as the process by which such organic molecules may be constructed from carbon dioxide, water, and minerals in the soil (Taylor et al., 2012).

Complete and correct understanding of carbon cycling as measured by the Grandma Johnson question, or others like it, would be difficult to attain by attending to pairwise interactions or even by attending to the complex molecular interactions that occur within mitochondria and chloroplasts (such as the set of chemical reactions in glycolysis). Yet, by attending to linked processes of photosynthesis, cellular respiration, and decomposition, big picture understanding appropriate to an introductory biology course can emerge.

METHODS

To determine whether students who use the newly developed LSET curriculum gain the desired big picture understanding of central life sciences processes, we assessed our students’ responses on the Grandma Johnson question (Ebert-May et al., 2003). Four groups of students were asked the Grandma Johnson question on a final exam.

High school: Students in grades 9–12 who either took a yearlong biology (n = 26) or advanced placement (AP) environmental science (n = 20) course in high schools in Bellevue, Washington. These high school courses could be characterized as typical high school science courses that adopt a fairly traditional, teacher-centered, pedagogical approach with a combination of whole-class lecture and small-group labs. The biology course was required of all 9th grade students for graduation. All students enrolled in the AP environmental science course were 11th or 12th graders.

College: Community college students from Everett Community College, Everett, Washington, who took either a quarter-long introductory biology for nonmajors (n = 23) or an introductory environmental science (n = 11). These college courses could be characterized as typical college, nonmajors science courses, although the environmental science class would typically have more of an emphasis on matter and energy transfer than would the general biology class. The two groups of students were very similar in terms of their background (all were nonscience majors.) We pooled their data because the number of students enrolled in the environmental science course was low, the students were all from the same institution, both courses were for nonmajors, and their data were very similar.

Quarter LSET 1.0: Community college students from Everett Community College, Everett, Washington, who took the original quarter-long version of the LSET curriculum (n = 26). Students in this class are not science majors and have minimal science background. Approximately 33–50% of the students in the class plan to become elementary school teachers. These data were collected early in the development of the curriculum, prior to adding an additional 5 wk of curriculum to address cell biology and genetics. However, the bulk of the questions, discussions, and activities relevant to cellular respiration, photosynthesis, decomposition, and energy and matter cycling were completed by this time.

Semester LSET 2.0: College students from CSU, Chico who took the semester-long version of the LSET curriculum (n = 45). Students in this class are not science majors and typically have minimal science background. Approximately 95% of the students intend to become K–8 teachers. These data were collected late in the development of the curriculum and are representative of the revised version being disseminated currently.
The results were scored according to the coding scheme shown in Figure 6A, which was derived from a grounded coding of student responses. Each student’s response was evaluated for the presence or absence of major target biological processes (A: decomposers consume Grandma Johnson’s remains and perform cellular respiration, releasing carbon dioxide into the environment; B: a plant takes in carbon dioxide and makes glucose through photosynthesis; C: an herbivore consumes the plant; D: the coyote consumes another organism) and those of common misconceptions (W: the coyote consumes a plant; X: a plant absorbs matter from Grandma Johnson’s remains through its roots; Y: worms or other detritivores consume Grandma Johnson’s remains; Z: the coyote consumes Grandma Johnson’s remains). A blank, vague, or nonsensical answer was coded N. If a pathway was omitted, it was not given a code as part of either a correct or incorrect answer. See Table 3 for example student responses and a brief description of how each was coded. The two of us responsible for coding the answers communicated closely on the coding scheme. Interrater reliability was established by
Table 3. Sample student responses and codes to the “Grandma Johnson” question

<table>
<thead>
<tr>
<th>Example student response</th>
<th>Score</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Her body would have its energy absorbed into the ground and then that energy is moved</td>
<td>N</td>
<td>This vague and nonsensical answer confuses energy with matter.</td>
</tr>
<tr>
<td>from the ground into the coyote’s leg because of recycled energy and recycled matter.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grandma Johnson decomposed into various nutrients which entered into the soil of the</td>
<td>XCD</td>
<td>This student thinks that decomposed matter enters plants through their</td>
</tr>
<tr>
<td>ecosystem. These nutrients were used by plants to grow. These plants were then eaten</td>
<td></td>
<td>soil.</td>
</tr>
<tr>
<td>by a consumer, such as a rabbit. That rabbit was then consumed by the coyote.</td>
<td>ADY</td>
<td>This student knows that a decomposer can perform cell respiration,</td>
</tr>
<tr>
<td>The decomposer would eat her carbs and then the decomposer would gain carbs.</td>
<td></td>
<td>however, misses the role of plants.</td>
</tr>
<tr>
<td>Through cellular respiration by the decomposer carbon would be released into the</td>
<td>BCD</td>
<td>This incomplete answer fails to account for how carbon dioxide was</td>
</tr>
<tr>
<td>environment. The coyote would then eat a decomposer which has carbs and now the</td>
<td></td>
<td>generated. Because a decomposer was not mentioned, that pathway (A)</td>
</tr>
<tr>
<td>coyote has its own carbs.</td>
<td></td>
<td>was not coded.</td>
</tr>
<tr>
<td>Her matter is used by the plant (CO₂ for example), which is eaten by the herbivores,</td>
<td>ABCD</td>
<td>This example of correct understanding illustrates a complete and</td>
</tr>
<tr>
<td>which is eaten by the coyote. This matter is used to perform certain tasks, and in this</td>
<td></td>
<td>accurate account of the carbon cycling in this scenario.</td>
</tr>
<tr>
<td>case performing a muscle leg of a coyote.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Grandma Johnson (now a dead organism) would store the carbon atom as a carbohydrate;    
when she is eaten by bacteria (a decomposer) the carbon atom will transfer to the       
bacteria as a carbohydrate. After cellular respiration the carbon atom will be released 
into the atmosphere as a CO₂ molecule. This CO₂ molecule will be used by a plant 
during photosynthesis and will be stored as a starch within the plant. The plant will 
then be eaten by a mouse who will store the carbon atom in the form as [sic] glyco-    
gen. When the mouse is eaten by a coyote the carbon atom will be transferred to a muscle  
cell in the coyote’s leg.

both of us coding 46 of the same student responses and comparing codes. Interrater reliability was 93.5%, with a Cohen’s kappa statistic of 0.73. 
In addition to a qualitative assessment of these data, we conducted an independent samples Kruskal-Wallis test to determine the level of differences between the response distributions across the four courses (high school, college, quarter LSET 1.0, and semester LSET 2.0) and whether there were any differences between the percentage of students that identify a correct pathway (ABCD) versus the most common incorrect pathway (XCD), all other incorrect pathways, or nonanswers (N) in each group.

RESULTS

The analysis of the Grandma Johnson data showed clear differences between the responses of students who were and were not exposed to the LSET curriculum (see Figure 6, B–E).

High school students completing an introductory biology course or an AP environmental science course showed little understanding of carbon cycling. Twenty-two percent of the students provided vague, incoherent responses, such as: “I don’t know” and “The carbon atom is broken down and Grandma Johnson eventually decomposes.” Only one student was able to identify the complete correct pathway. 
College students in introductory biology and introductory environmental science were less uncertain (only 3% were scored N), but, on the whole, gave a surprising number of incorrect answers. Seventy-one percent of these students assumed that the plant absorbed carbon atoms from Grandma’s remains through their roots.

Performance on this question improved dramatically with the LSET curriculum. Forty-six percent of students taking the early, quarter-long version of the course and 64% of students taking the revised, semester-long version of the course were able to identify the complete correct pathway.

An independent samples Kruskal-Wallis test showed that the distribution of student responses (correct, incorrect [XCD], incorrect [all others], or blank/vague) across the four courses were not statistically similar (p < 0.01 level). In other words, the distribution of responses changed significantly as students matured (between high school and college) and between students who took different versions of the LSET curriculum (between college, quarter LSET 1.0, and semester LSET 2.0). These data are represented in Figure 7.

DISCUSSION

As we strove to create an exemplary life sciences curriculum modeled after PET, we anticipated that the format would adapt well to biological content; the PET curriculum is based on learning principles that are largely independent of content, and we therefore expected that many aspects of the
curricular design would transport easily and well. For example, both curricula place great emphasis on eliciting, acknowledging, and explicitly addressing students’ commonly held initial ideas. Similarly, the daily structure of PET lessons with their emphasis on cooperative learning and hands-on experimentation flowed readily into the life sciences context.

However, there were also considerable challenges to overcome—challenges that would likely impact any effort to adapt proven curricula from physics education to fit the needs of biology at the introductory level. Primary among these considerations is a difference in whether the curriculum emphasizes pairwise interactions or linked processes involving many interactions. Whereas introductory physical science presents itself as offering explanations and mechanisms at the level of individual interactions between objects and molecules, introductory biology considers systems such as living organisms and ecosystems that are far more complex. Introductory biology does not attempt to offer explanations at the level of pairwise interactions—there are simply too many to consider for a nonscience major. Rather, understanding processes at the big picture level of organisms and ecosystems becomes a far more relevant and important goal.

The consequences of targeting ideas at the process level is that many features of well-developed physics curricula such as PET required significant reworking to find success in a life sciences context. Representations of energy transfers cannot represent pairwise interactions—rather, biology must focus on processes. Definitions cannot be presented as a simple, straightforward given—constructing definitions becomes part of the scientific inquiry, embedded within an understanding of significant biological processes. Anomalous experimental results cannot simply be attributed to experimental error—live organisms, with innumerable interactions underlying the processes we investigate, do not always behave in anticipated ways, and isolating particular interactions is neither feasible nor does it add core introductory biology topics. And scientists’ ideas cannot wait until the end of a chapter—some biological ideas need introduction and cannot be constructed from first principles in the introductory course.

Successfully addressing these challenges affords superior student outcomes in that students may then fully explore and come to understand central biology concepts from a big picture perspective. After completing our LSET curriculum, our students can integrate and apply concepts from across life sciences topic areas (such as photosynthesis, decomposition, and ecology) better than high school or college students who completed more traditional biology and environmental science curricula, as we found with the results of the Grandma Johnson question. Students who had taken LSET had a better understanding of the core concepts of the flow of matter and energy in living systems and could relate them to the fundamental principles of the conservation of energy and matter.

This is partly due to the pedagogy of the course. As is advocated in best practices of biology education (and science education in general), students are afforded time to wrestle with ideas and to construct understanding. Prior knowledge is elicited and explicitly addressed. Students are prompted to record how their thinking has changed. However, we think the success of our students is also due to how our course is thematically and pedagogically linked with PET. Linking PET and LSET allows students to better use principle-based reasoning, especially in terms of the flow of energy and matter. PET explicitly covers the law of conservation of energy in the use of energy diagrams. By adapting energy diagrams into our curriculum, and by adding matter diagrams, we enabled students’ exploration of the connections between these laws of physics and the processes of biology.

The curriculum also allows students to learn in a manner that addresses some of the reasons elementary teachers feel unprepared to teach science. Rather than presenting biology as a collection of facts to be memorized, LSET presents it as a way to understand the living world. Students come to understand that biology is about asking and answering questions. This process is modeled for them throughout the course. Additionally, the Learning about Learning activities explicitly link what they will experience in the curriculum in their future classrooms.

Physics has a long tradition of research-based curricula at the undergraduate level, with methods that may be leveraged for use in developing curricula for undergraduate biology. However, leveraging these approaches requires more than importing biology content into formats established in physics. Attention to the ways in which pedagogical innovations support particular disciplinary norms and how those norms differ between disciplines is critical for adapting innovations into new fields.

ACKNOWLEDGMENTS

We are grateful to the many students who have been in our classes and have given us helpful feedback on this curriculum and to our colleagues who were part of the development team. This work was funded through NSF grants 0313060 and 0942391. The manuscript was greatly improved by thoughtful comments from two anonymous reviewers.

REFERENCES


Article

Addressing the Challenge of Diversity in the Graduate Ranks: Good Practices Yield Good Outcomes

Nancy L. Thompson*† and Andrew G. Campbell†‡

*Office of Graduate and Postdoctoral Studies, †Brown University Initiative to Maximize Student Development Program, Division of Biology and Medicine, and ‡Department of Molecular Microbiology and Immunology, Brown University, Providence, RI 02912

Submitted April 30, 2012; Revised October 15, 2012; Accepted October 24, 2012

Monitoring Editor: Jose Vazquez

In this paper, we examine the impact of implementing three systemic practices on the diversity and institutional culture in biomedical and public health PhD training at Brown University. We hypothesized that these practices, designed as part of the National Institutes of Health–funded Initiative to Maximize Student Development (IMSD) program in the Division of Biology and Medicine, would have a positive effect on underrepresented minority (URM) recruitment and retention and objective measures of student success. These practices include: 1) develop strategic partnerships with selected undergraduate institutions; 2) provide a personalized education program of student support and skill-based modules to supplement discipline-based course work; and 3) transform institutional culture by engaging faculty in supporting diversity-related goals and practices. Data comparing URM numbers and key academic milestones before and after implementation of IMSD practices support the initial hypothesis and effectiveness of these practices at Brown. Program components are broadly applicable as best practices for others seeking to improve URM recruitment and achievements of graduate students traditionally underrepresented in the sciences.

INTRODUCTION

The identities and missions of academic institutions are shaped by practices that impact the eventual makeup of the communities. Commitment to and successful implementation of practices that achieve diversity, for example, will translate into communities whose populations mirror the U.S. population as a whole. One area in which progress toward diversity has been slower than desired has been in graduate education in the biomedical, life, and public health sciences (Chang et al., 2008; Aud et al., 2010). Racial and ethnic minorities account for 27.9% of the U.S. population, yet they are among the most underutilized and underrepresented (UR) groups in the scientific workforce on a proportional basis (National Academy of Sciences, 2011; National Science Foundation [NSF], 2012, Table 1 and Table 7-4). Blacks, Hispanics, and Native Americans account for only 7.1% of all employed biological/biomedical and life scientists (National Academy of Sciences, 2011; NSF, 2012, Table 9-6). Cutting across race and ethnicity is low or poor socioeconomic status, both of which influence career choices and opportunities (Leppel et al., 2001; Ward, 2006). While efforts continue to be made to recruit underrepresented (UR) individuals to disciplines in the biological and biomedical sciences, challenges of access, motivation, retention, and academic and social support persist (National Academy of Sciences, 2011). These challenges exist well beyond training stages. Recent data show that, despite all other things being equal, African-American scientists are 10 percentage points less likely to be funded by the federal peer review–driven National Institutes of Health (NIH) R01 funding mechanism relative to their majority colleagues (Ginther et al., 2011). More broadly, underrepresented minorities (URMs) comprised just 6.5% of the NSF-funded investigators (National Academy of Sciences, 2011).

Many factors have been proposed to explain the generally poor outcomes of recruitment, retention, and success of UR
individuals especially ethnic and racial minorities, yet no one unifying solution has been reached that adequately addresses the problem. Program development must recognize that lack of diversity and representation may not simply reflect lack of preparation, but also exists as an opportunity to identify and address institutional weaknesses that lead to underrepresentation and poor success.

We sought to identify and address issues affecting diversity in the nine BioMed PhD-training programs at Brown. The BioMed division does not admit students for graduate studies through an “umbrella” program. Instead, it requires that applicants preidentify a disciplinary program in biology or public health at the time of application. Each program therefore develops its own recruitment strategies and builds its own applicant pool. Despite an overall 27% population growth from 193 students to 247 students from 2004 to 2007, only five of the nine BioMed PhD-training programs during that period enrolled students who identified as URM students. In 2006–2007, we examined key practices utilized by the highly diverse Pathobiology program and asked whether these could be implemented more widely as part of an Initiative to Maximize Student Development (IMSD) program, and whether they would translate to increased diversity and student achievement across biology and public health. In doing so, three broad questions were investigated:

1) Does development of institutional partnerships enhance URM recruitment and student success? 2) Does addressing gaps in undergraduate preparation by creating a system of personalized student support, enhance student academic success? 3) Does increased faculty involvement in interventions and shared goals impact institutional culture and diversity outcomes?

We present data on measurable outcomes assessed before and after IMSD practices were broadly implemented. Though challenges remain in causal interpretation, overall student diversity and several accepted indicators of academic achievement increased significantly in the 3 yr after IMSD was established. Furthermore, faculty from all nine graduate programs in Brown’s Division of Biology and Medicine are now involved in the interventions described. This report describes a series of pre-emptive academic and nonacademic steps designed to maximize UR student success, but also of benefit to all students. The steps are broadly applicable to other institutions and programs.

METHODS

Data Analysis

Percent URM student enrollment for Pathobiology was compared with overall data for the Division of Biology and Medicine and national figures for diversity in biomedical PhD programs. Data on GRE scores and GPA for matriculating URM students versus all students were also examined. Data for the 2008–2011 academic years are referred to as the “IMSD era.” Data from 2005–2007, the “pre-IMSD era,” represent baseline data used for comparative analysis.

We report on the 3-yr period after the IMSD grant began to provide information on early outcomes of graduate matriculants, including IMSD trainees on the path to the PhD. This represents a reasonable period for measurable outcomes to be achieved without the additional time period needed to reach awarding of the PhD. For matched-cohort analysis of measures including student publications, presentations, federal or national predoctoral fellowships, and nonminority travel awards, only students who had completed at least 1 yr of graduate studies were included in the calculations. Excluded from these calculations is the Pathobiology graduate program on which the Brown IMSD program was based. All IMSD trainees were URM students. IMSD-era calculations were performed by comparing 17 IMSD-supported students with a group of 17 non-URM students matched by graduate programs and entry year to PhD training. For the pre-IMSD era calculations, URM trainees were compared with non-URM trainees matched by graduate program and year of entry.

Build Institutional Partnerships

Criteria and Goals. IMSD institutional partner relationships were developed as byproducts of relationships established between faculty members at each of the participating institutions with the former director of the Pathobiology PhD program and were sustained as he became codirector of the IMSD program. These relationships were guided by common goals of increasing URM enrollment in biomedical and public health PhD programs and achieving measurable outcomes of trainee success. Though recognized at the institutional administrative level, the partnerships were faculty-driven, and relatively informal relationships were created in a manner similar to research-based scientific collaborations administered by faculty. Criteria used in seeking partnerships were high URM undergraduate enrollment, quality students interested in research, and a relatively easy travel distance from Brown.

Activities. A regular schedule of visits by the IMSD codirector to the partner institutions and partner leaders to Brown was established and took place with a specific agenda for each visit. Among the important activities driven by the relationships is a practice of “curricular mapping” or sharing and exchanging of details regarding curricular content, content delivery, and means of assessment. We hypothesized that shared information would improve teaching outcomes and student readiness for graduate work. The mapping process looks at alignment of course syllabi and learning methodology between key undergraduate courses and first-year graduate academic expectations. Faculty share information during campus visits and partner faculty receive feedback from their alumni who matriculate to Brown. Gaps identified in preparation of students matriculating to Brown are addressed early in advising, with referral of students for enrollment in IMSD-sponsored, supplemental skill-based training modules (described in Personalized Student Support).

Assessment. Partner leaders are known to each other and are also members of an external advisory group to IMSD. An IMSD external evaluation consultant hired as part of the IMSD grant periodically contacts institutional leaders to assess progress and challenges in expanding URM entry and student success in biomedical and public health PhD programs. Recommendations based on the assessment are prepared and presented to IMSD leadership.
Personalized Student Support
Skill-Based Training Modules. A menu of noncredit, active-learning experiences was developed to fill preparatory gaps identified either as a result of curricular mapping, student self-referral, or faculty advising. The “module” topics are based on competencies widely acknowledged to be critical to graduate student success. Their overall goal is to provide essential foundational knowledge that maximizes student readiness for various phases of graduate training. Learning objectives and assessment criteria for each were developed with significant input from the Brown Sheridan Center for Teaching and Learning.1 Each module consists of intensive training sessions of 10–12 contact hours, offered over 1–2 wk in a classroom or laboratory setting, depending on the content delivered. IMSD-supported trainees are required to complete a minimum of three modules selected in consultation with the IMSD director and faculty advisors. Other graduate students who enroll in modules include those who self-select or are referred by their faculty advisor or graduate program director. Enrollment in each module is capped at 15 participants and enrollment priority is granted to IMSD trainees. Modules are coparticipated by faculty members and an advanced graduate student who serves as a peer mentor with the title “senior scholar.” The majority of modules are offered during the academic year and run concurrent with regular graduate courses.

Advising and Tracking of IMSD Student Progress. The qualifications for “IMSD students” (trainees supported by IMSD funds during any time in their graduate training) include status as matriculated UR students identified as those who would benefit from IMSD resources and training. IMSD student appointees must also meet several expectations of the IMSD program that exceed the expectations for other graduate students. These include involvement in ancillary activities that support their academic development and meet with the approval of each student’s graduate program director. Activities include participation in skill-based training modules, peer mentoring, and submission of progress updates, as described later in Results of this report. Trainees’ progress is closely monitored on a quarterly basis and during regularly scheduled annual advisory meetings held jointly with IMSD directors and the students’ graduate programs. In addition to tracking academic performance in the classroom, close monitoring of research progress occurs, as trainees are expected to submit written summaries upon completion of each research rotation prior to selecting a permanent lab for thesis work. Students who have already selected their thesis lab are required to complete research summary reports once a year for the life of their graduate careers. Because challenges are posed by the unique curriculum and mode of research training offered by each graduate program, especially when field-based research is used rather than bench-based research, a common format was developed in which each IMSD student provides the title of his or her research project, a clearly stated driving hypothesis, and the methodology. When appropriate, students state progress made since their last progress report submissions and outline plans for the coming year. Although written by the students, these reports are reviewed and endorsed by their research or thesis advisors. Early in training, these reports are used to evaluate student academic development, communication skills, and progress toward the preliminary examination. Trainees are expected to complete the preliminary examination no later than the end of the summer that marks the end of their second year as graduate students. Beyond the second year, trainees submit reports that are used to monitor their progress toward thesis completion and the PhD degree defense. Although the Brown IMSD program activities support UR students, many of these activities and program resources are accessible to all other matriculated graduate students.

Change the Culture of the Institution: Faculty Involvement in IMSD
Faculty perceptions and expectations impact admissions practices and resulting student diversity within graduate programs. We sought to expand diversity and change the institutional culture by engaging faculty on several levels through IMSD program activities. The directors of the nine PhD programs within the Division of Biology and Medicine were appointed to the IMSD Internal Advisory Committee and asked to make recommendations of incoming students for IMSD student support, to refer matriculated students for skill-based training modules, and to serve as faculty leaders of training modules or to suggest faculty colleagues who would be effective in this role. Support for selected UR students via IMSD funds was offered to encourage consideration of a wider range of applicants than might have been considered in the past, with the benefits that participating graduate programs could expand in size and potentially increase their competitiveness for NIH T32 training grants.

RESULTS
Changes in URM Student Percentages
Data on diversity for Brown BioMed relative to national statistics are shown in Figure 1. The Pathobiology graduate program served as a model program for IMSD. In Pathobiology, ∼30% student diversity was achieved over a 5-yr period, as shown in Figure 1. At the program’s peak, 17 of the 51 enrolled PhD students were URMs. Attrition of UR students has been consistently low in Pathobiology, well under 5%, indicating that admitted students were able to meet program requirements for the PhD. Total numbers of URM graduate students across the BioMed division are shown as a percentage of the total student population for the 2005–2006 through 2010–2011 academic years (Figure 1). The total BioMed student populations for this same period were 237, 247, 256, 272, and 294, respectively. A steady increase in the percentage of UR students has taken place, with a peak of 21% in 2009–2010 (Figure 1). Nationally, UR enrollment for the same period was ∼10%. From the 2005–2006 through 2010–2011 academic years, the BioMed graduate student population grew from 237 students to 289 PhD students. During this period, the numbers of URM students in the life and public health sciences increased from a low of seven students to a high of 57.

---

1The Harriet W. Sheridan Center for Teaching and Learning (http://brown.edu/Administration/Sheridan_Center) provides and supports a range of pedagogical approaches to teaching and offers support to all members of Brown’s teaching community. The center also recognizes the diversity of learning styles and uses a number of mechanisms to encourage reflective, independent, lifelong learning on the Brown campus.
The increase coincided with renewed institutional commitment and interinstitutional relationships forged by the Brown IMSD program. Although the numbers of URM candidates admitted to BioMed graduate programs has increased since the IMSD program began in 2008, a drop in the total percentages and actual numbers of URM students among the graduate student population has taken place in the most recent 2 yr. This reflects the fact that previously admitted URM students, most of whom were in the Pathobiology program, have graduated at a rate faster than new URM students have matriculated across the division.

Institutional Partnerships to Enhance URM Recruitment and Student Access

Brown IMSD established partnerships with institutions serving significant numbers of UR students. The current institutional partners include: 1) York College, a small public college that is part of the City University of New York; 2) the Queens, New York, campus of St. John’s University, a relatively large multicampus private university; and 3) North Carolina A&T State University, an HBCU (historically black college or university) that is part of the multiinstitutional University of North Carolina system. All of these campuses have highly diverse undergraduate student populations. York College serves a population primarily of first-generation, college-going African-American and Hispanic students and was ranked 10th among the top 50 non-HBCUs awarding bachelor’s degrees to black Americans. The Queens campus of St. John’s University serves a broad constituent of racially and ethnically disadvantaged students. North Carolina A&T serves a large URM student population who are primarily southern U.S. African Americans. These partner institutions are unique in that they do not represent the traditional high-profile HBCUs or Hispanic-serving institutions often viewed as the bellwethers of academic accomplishments. Our partners are, however, similar to these institutions in sharing the same mission of providing all students, regardless of race and background, the opportunity to develop their academic skills and talents en route to joining the domestic private and public U.S. workforce.

Activities that take place as part of the partner relationships are listed in Table 1. The partner relationships provide frequent opportunities to meet with and engage students from partner institutions at regional and national scientific conferences, as well as at their home institutions. The scope of interactions extends beyond the admissions season. The partnerships have facilitated connections via personal visits with undergraduates and master’s-level students at the partner institutions and groups of students at early stages in their training at Brown. As a result, the Brown IMSD program has been able to make students better aware of educational and career options available to them beyond their local communities. To sustain interinstitutional ties, Brown IMSD supports Brown faculty visits to partner institutions. In these settings, faculty engage students about research areas of interest, career aspirations, and expectations. Faculty also use these opportunities to demystify the PhD-training process and to introduce career options students may not have considered. These personalized engagements help to diminish persistent and widespread misconceptions about graduate school, such as the range of science career options available to PhDs and the financial costs associated with earning the degree. Over the past 3 yr, Brown IMSD has taken an active role in the curricular mapping process with partner institutions to improve pedagogical outcomes. The mapping process has helped to raise awareness about areas of weakness common to beginning graduate students, particularly inquiry analysis and problem-based learning. Faculty visits and regular communications developed as part of Brown IMSD have permitted the type of open exchanges with partner institutions crucial to this process.

Changing Applicant Pool Diversity

IMSD’s ability to build cross-institutional connections with students at partner institutions has resulted in an increase in student applicants and matriculants to Brown graduate programs from these partner institutions. Since the start of the partner relationship, student representation from the partner institutions has gone from 0% of the total PhD student population to 3% of the total graduate student population. The start of the Brown IMSD program in 2008 heralded an aggressive and systematic effort to attract a diverse cohort of graduate students across Brown’s nine graduate programs in biomedical and public health sciences. As shown in Figure 2, URM students accounted for 23% of all matriculating PhD students admitted to BioMed graduate programs in 2011–2012, compared with 17% during 2007–2008, the immediate pre-IMSD period. The rate of increase in the numbers of UR graduate applicants during the IMSD era, especially URM students, exceeded the rate of increase of non-URM applicants seeking admissions during the same period. For the 2011–2012 academic year, URM applicants accounted for 14% of all U.S. applications, compared with 11% during the most recent year of the pre-IMSD era (2007–2008).

Admitted-Student Qualifications

Aggregated data regarding matriculant credentials were evaluated as one measure of student quality. These analyses include grade point averages (GPAs) and Graduate Record
Table 1. Brown IMSD program partner institution activities: inter-institutional partnership components

<table>
<thead>
<tr>
<th>Activity</th>
<th>Goal</th>
<th>Frequency</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduate curricular</td>
<td>Exchange of academic content aligning undergraduate and graduate curricula to improve student readiness for PhD studies and success</td>
<td>Semiannually</td>
<td>Meet, exchange, and share pedagogy with partner institution faculty. Current graduate students share experienced weaknesses and strengths of undergraduate preparation with their undergraduate institutions</td>
</tr>
<tr>
<td>mapping</td>
<td></td>
<td></td>
<td>Faculty representatives from each institution spend 2 d at Brown, meeting with faculty (including graduate program directors), students, and administrators and attending classes and seminars.</td>
</tr>
<tr>
<td>Partners’ meeting</td>
<td>To advance work on enhancing and strengthening strategic partnerships and transforming culture of diversity</td>
<td>Annually</td>
<td>Faculty representatives from partner institutions informally engage IMSD trainees, IMSD faculty, IMSD senior scholars, and administrators.</td>
</tr>
<tr>
<td>IMSD program launch</td>
<td>Build stronger interinstitutional connections; build faculty bridges; assist partner institution faculty in monitoring the point-to-point maturation of their past students</td>
<td>Annually</td>
<td>IMSD invites guests from partner institutions for a 1 1/2-d visit to Brown to meet their past alums and other graduate students, attend seminars and classes, and visit research labs. Support for the visit is provided by the associate provost and director of institutional diversity.</td>
</tr>
<tr>
<td>Partners’ Day</td>
<td>Provide partner institution students and faculty with firsthand experience on the Brown academic environment and its training climate for graduate students in the life, biomedical, and public health sciences</td>
<td>Annually</td>
<td>IMSD program director identifies students declaring interest in science and tracks them starting in their freshman year. Students meet with IMSD faculty on each visit to partner campus.</td>
</tr>
<tr>
<td>Student relationship</td>
<td>Build deeper and more lasting relationships with URM students starting in their freshman college year; maximize student opportunities for pursuit and success in science.</td>
<td>Annually</td>
<td>IMSD program hosts an informal social gathering of partner institutions’ students and faculty.</td>
</tr>
<tr>
<td>building</td>
<td></td>
<td></td>
<td>IMSD program hosts an informal social gathering of partner institutions’ students and faculty.</td>
</tr>
<tr>
<td>IMSD hosted socials</td>
<td>Cultivate and strengthen partner relationship; extend relationship beyond the admissions season</td>
<td>Annually</td>
<td>IMSD program hosts an informal social gathering of partner institutions’ students and faculty.</td>
</tr>
</tbody>
</table>

*Annual and semiannual activities of the Brown IMSD partners are outlined. Information presented in the table represents shared interests, activities, and expectations of all partners and serves as a working guide for interinstitutional relationship building.*

Figure 2. Comparison of URM student applicants, admitted students, and matriculants in BioMed graduate programs before and after implementation of the IMSD program. The figure shows URM students as a percentage of all PhD graduate students. Left, URM students applying to, admitted to, and matriculating in, respectively, PhD studies in 2007–2008, the year immediately preceding the start of the Brown IMSD program. Right, URM students applying to, admitted to, and matriculating in, respectively, PhD studies in 2011–2012, the most recent year of the IMSD era. Values are for racial and ethnic minorities and exclude other UR students applying to, admitted to, and matriculating in BioMed graduate programs.

Examination (GRE) scores (verbal and quantitative), which are summarized in Figure 3. Changes in GPAs of applicants to BioMed graduate programs since the start of the Brown IMSD program were evaluated, as shown in Figure 3a. Among the students admitted in 2011–2012, there was an expansion at both the high and low ends of the GPA ranges for all matriculating students. Despite the broadening of the GPA range among admitted students, the mean GPA for admitted students was 3.42 in 2011–2012, compared with 3.29 in 2007–2008. The low end of the GPA range of admitted URM students during the IMSD era did not differ from the value for URM students admitted in the pre-IMSD era. URM students applying to, admitted to, and matriculating in BioMed graduate programs during the IMSD era, however, no longer define the low end of the GPA range of applicants, admitted students, or matriculants. These changes reflect positively on our diversity efforts and are in line with the goals of the IMSD program.

GRE verbal score data are shown in Figure 3b. Verbal scores for URM students matriculating in 2007–2008 ranged from 410 to 710 among a total matriculating pool with a range of 320–760. This compares with the 2011–2012 URM matriculants, who had a verbal score range of 410–690 in a total pool with a range of 350–700. The verbal score range over which non-UR students perform is thus broad and encompasses the score range for UR students. The 2011–2012 GRE verbal score ranges for both
Addressing the Challenge of Diversity

Figure 3. Box-and-whisker plots of (a) undergraduate GPAs and (b) verbal and (c) quantitative GRE scores of matriculants to PhD-training programs of BioMed. GPA and GRE scores are presented for admitted URM (shaded boxes) and non-URM matriculants (open boxes) for 2007–2008 (pre-IMSD) and 2011–2012 (the most recent IMSD year). (a) The mean GPA for students admitted in 2007–2008 was 3.29 vs. 3.42 for students admitted in 2011–2012. (b) Verbal GRE scores for URM matriculating in 2007–2008 ranged from 410 to 710 among a total matriculating pool with a range of 320–760. The 2011–2012 URM matriculants had a verbal score range of 410–690 in a total pool range of 350–700. (c) In the 2007–2008 academic year, URM matriculant quantitative GRE scores ranged from 470 to 750 among a matriculant pool whose quantitative scores ranged from 470 to 800. URM students admitted in 2011–2012 had a quantitative score range of 410–760 among a matriculating pool with a range of 410–800.

URM and non-URM matriculants has narrowed, with much of the change in range taking place at the higher end of the range.

Data for the GRE quantitative exam are shown in Figure 3c. For the 2007–2008 academic year, scores ranged from 470 to 750 for URM matriculants among a matriculant pool whose quantitative scores ranged from 470 to 800. In comparison, URM students admitted in 2011–2012 had a quantitative score range of 410–760 among a matriculating pool with a range from 470 to 800. For 2011–2012 URM and non-URM matriculants, the low end of the GRE quantitative score range was markedly lower than for matriculants admitted during 2007–2008, the most recent pre–IMSD era admissions period. This change likely reflects the increased confidence Brown graduate programs place in the ability of Brown IMSD to address weaknesses in the quantitative skills of students who begin their graduate training. There is also only a negligible increase at the high end of this range. Although the quantitative score ranges have broadened, the average score for all matriculants increased from 620 for 2007–2008 matriculants to 645 for 2011–2012 matriculants. URM matriculants were not the sole determinants of the low end of the quantitative score range of matriculating students.

URM applicants did not define the low end of the ranges in either the applicant or accepted applicant pools for the verbal, the quantitative, or the analytical (unpublished data) components of reported GRE scores. GRE test scores in general vary widely among graduate applicants, and they are often difficult to correlate with student abilities and future success in graduate school (Morrison and Morrison, 1995; FairTest, 2007). Anecdotally, this has been found to be the case for...
many past applicants to our graduate programs, regardless of their background.

Despite the changes observed in score ranges, analysis of adjusted p values using the Benjamini and Hochberg method for URMs admitted during the pre-IMSD era versus URMs admitted during the IMSD era shows that there is no statistical difference in the average GPAs and average GRE scores of admitted students. This analysis, however, does not fully reveal the details of the students admitted. A careful examination of the average GPAs and GREs of students admitted in the IMSD era shows that the absence of statistical change is the result of graduate programs matriculating students that have numerical GPA and GRE scores higher than students admitted in the pre-IMSD era, as well as simultaneously matriculating students with scores weaker than scores of students admitted in the pre-IMSD era. While GRE scores may not be useful predictors of student success, they may be more useful as diagnostic tools that identify gaps that need to be addressed. Students admitted to graduate studies who performed poorly on one part of the GRE are given supplemental module training in their first year at Brown. Specifically, graduate students performing poorly on the quantitative portion of the test are enrolled in IMSD skill-based modules designed to strengthen their quantitative skills. Similarly, students performing poorly in writing enroll in modules that strengthen writing skills.

**Personalized Student Support to Enhance Student Academic Success**

Academic and research knowledge gaps of students beginning graduate training at Brown are addressed using 11 not-for-credit, skill-based training modules developed by the Brown IMSD program. The active-learning process of each module provides structured training exercises as preludes to formal course work and other training experiences. The current menu of modules, learning objectives, and competencies addressed is listed in Table 2 and the Supplemental Material. Although initially designed to address the needs of UR students, the value of these modules has become widely accepted as useful training tools for all graduate students. Each year, more than 20% of the BioMed graduate student population takes at least one module, and ~74% of all module subscribers are not IMSD-supported graduate students. Additional module subscribers include visiting faculty and postdoctoral fellows. Anonymous surveys conducted at the conclusions of modules reveal a high degree of satisfaction with the material covered and its perceived benefit to participants (unpublished data). The high level of enrollment by all students and involvement of faculty from all BioMed graduate programs as module leaders have pre-empted the perception of modules as “remedial” or otherwise associated with negative stigma.

**Milestones to Evaluate Progress**

During the IMSD era, all nine BioMed graduate programs have now been successful in matriculating UR students (Figure 4). Given the challenges of admitting UR students to programs in which they are underrepresented, this success reflects the effectiveness of institutionalizing good practices.

The Brown IMSD program is entering its fourth year of operation, and the number of trainees who have advanced to the preliminary examination or thesis defense is therefore too low to utilize as a measure of IMSD success. Several measurable milestones prior to the preliminary examination are used to assess student trajectory and simultaneously evaluate the effectiveness of IMSD practices. These milestones have been particularly useful in early assessment of the achievements of IMSD-supported students relative to their peers and in

---

**Table 2.** Skill-based training modules developed and administered by the Brown IMSD program

<table>
<thead>
<tr>
<th>Module</th>
<th>Content or purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demystifying the PhD Experience: Strategies for Academic &amp; Personal Success in Grad School</td>
<td>Become aware of and develop strategies to implement and fully integrate the academic and nonacademic skills required to succeed in graduate school</td>
</tr>
<tr>
<td>Beyond the Hypothesis: Experimental Design &amp; Critical Analysis</td>
<td>Develop skills in mechanistic hypothesis setting and experimental design</td>
</tr>
<tr>
<td>Designing and Delivering Scientific Presentations</td>
<td>Gain insight and practice in effective oral communications of scientific results</td>
</tr>
<tr>
<td>Defending Your Research Proposal &amp; Critiquing Those of Others</td>
<td>Strategies in selecting a strong thesis topic; evaluating your progress; giving and receiving advice</td>
</tr>
<tr>
<td>Resources, Tools and Basic Techniques in Molecular Biology</td>
<td>Gain insight into when to apply particular methods and resources for genomic/proteomic approaches</td>
</tr>
<tr>
<td>Professionalism: Maximizing Your Impact in Professional Settings</td>
<td>Recognize and acquire behaviors that promote career success in biology and public health</td>
</tr>
<tr>
<td>Scientific Presentation of Biological Data</td>
<td>Learn how to construct effective graphs that maximize meaningful content and interpretation</td>
</tr>
<tr>
<td>Scientific Writing: Key Principles</td>
<td>Learn strategies to effectively communicate in writing the what, why, how, and outcomes of your work</td>
</tr>
<tr>
<td>Introduction to Statistical Analysis of Data</td>
<td>Gain familiarity with statistical software and when to apply them in analyzing your data</td>
</tr>
<tr>
<td>Reading Scientific Publications</td>
<td>Develop skills in interpreting, critiquing, understanding, and appreciating journal articles in your field</td>
</tr>
<tr>
<td>Essential Laboratory Calculations</td>
<td>Pointers on accuracy, following protocols, and making measurements that are critical to experimental success and reproducibility</td>
</tr>
</tbody>
</table>

---

*Module subscribers include graduate students, postdoctoral fellows, and visiting and regular faculty. Enrollment number is capped at a maximum of 15 participants. Enrollment priority is granted to IMSD trainees, then non-IMSD graduate students. Each module is co-taught by a regular full-time faculty and advanced graduate student senior scholar.*
Addressing the Challenge of Diversity

![Graph: PhD Programs Enrollment Statistics]

**Figure 4.** Comparison of enrollment of URM PhD students among the nine BioMed graduate programs. The number of graduate programs making offers of admission to URM students in the pre-IMSD era (2007–2008) compared with programs making offers in the most recent year of the IMSD program (2011–2012). Also shown are the program success rates for being able to matriculate URM students. Matriculating students represent those to whom offers were made and who subsequently enrolled in graduate training at Brown.

Evaluating the readiness of these trainees for NIH T32 support. A comparison of the achievements of IMSD trainees with those of a group of non-IMSD trainees is summarized in Table 3. To avoid overrepresentation by some graduate programs that have student populations three to five times larger than other programs, Table 3 shows data for matched cohorts of students before and after implementation of the IMSD program. Measures of trainee achievements include publications on which trainees are listed as coauthors, securing of federal or national predoctoral fellowships, research presentations at scientific conferences, and receipt of nonminority travel awards. Calculations for the past 3 yr are based on data for six of the nine BioMed PhD programs, because these were the programs matriculating IMSD trainees during this period. The results show that IMSD-supported trainees publish at a rate greater than pre-IMSD era URM trainees and that they publish at a rate comparable with their contemporary non-URM peers.

IMSD-supported trainees were also awarded federal and other national fellowships at a rate greater than pre-IMSD era URM trainees. Although pre-IMSD era URM trainees received more travel awards than their non-URM, pre-IMSD era counterparts on a percentage basis, trainees supported during the IMSD era received more total travel awards than their pre-IMSD era counterparts. The absolute values in Table 3 presented for pre-IMSD era trainees are smaller than the values for the IMSD era, and the variables responsible for these differences are unknown and potentially complex. However, a comparison of ratios between URM and non-URM trainee achievements in 2005–2007 versus ratios between IMSD era–supported trainees and non-URMs in 2008–2011 shows that IMSD trainees perform at rates comparable with their non-URM peers during the IMSD era. This is in contrast to the performance of URM trainees relative to their non-URM peers in the period prior to the IMSD program. Only in terms of total travel awards and fellowships received did pre-IMSD era URM trainees do better than their IMSD-era peers. In terms of absolute values, however, IMSD-era URM trainees perform better than their pre-IMSD era counterparts. While the total number of trainees examined here is small, and the period of analysis is too short to define a clear pattern, the results reveal a trend that shows improvements in UR student performance correlating with IMSD program support and practices. These findings are also supported by statistical analysis. Fisher’s exact test shows that the increase in cohort publications among URM trainees during the IMSD era is significant, with a two-sided p value of 0.1314. The same type of analysis reveals that the increases in URM scientific presentations is also significant, with a two-sided p value of 0.0399. Despite increases in absolute numbers, however, there is no statistical significance in the increase of fellowships or travel grants awarded to URM students during the IMSD era for the cohort studied.

IMSD trainees are expected to demonstrate competency and progress in their graduate research and to develop their own research training plans. All are required to complete written quarterly or annual assessment of progress in graduate research competency, and these work products are used as foundations for preparing predoctoral fellowship

<table>
<thead>
<tr>
<th>Training period</th>
<th>Trainee status</th>
<th>Cohort publications</th>
<th>Fellowship awards</th>
<th>Scientific research presentations</th>
<th>Travel awards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre–IMSD era (2005–2007)</td>
<td>URM: No IMSD support</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Non-URM</td>
<td>4</td>
<td>1</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>IMSD era (2008–2011)</td>
<td>URM: IMSD support</td>
<td>15</td>
<td>6</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Non-URM</td>
<td>18</td>
<td>9</td>
<td>25</td>
<td>4</td>
</tr>
</tbody>
</table>

Measure of student achievement includes publications on which trainees are listed as coauthors, securing of federal or national predoctoral fellowships, research presentations at scientific conferences, and receipt of nonminority travel awards. Data are from six of the nine BioMed PhD programs (those that initially matriculated IMSD trainees). Data presented are for achievements over a 2-yr time frame.

For cohort publications, two-sided p value = 0.1314.

For scientific research presentations, two-sided p value = 0.03986.

Excludes minority travel awards.
applications for external support. This practice is shared widely with all graduate programs, as it is applicable to both URM and non-URM students. Figure 5 summarizes information gathered from NIH RePORTER (projectreporter.nih.gov/reporter.cfm). It shows that since the start of the IMSD era in 2008, a significant number of new NIH F-series predoctoral fellowships have been awarded to PhD trainees in BioMed. A number of these have been awarded to IMSD trainees. IMSD trainees and other students supported by the program have also secured additional fellowships from the Ford Foundation and the NSF at levels that exceed those previously secured by URM students in the pre-IMSD era (unpublished data). As shown in Figure 5, from 2008 to 2011, a total of 33 new F-series fellowships were awarded to Brown graduate students, and 10 of these were awarded to URM trainees. This is in addition to fellowships from NSF and other sources. The year 2011 also coincided with the period of “funding saturation,” such that the maximum number of students who could apply and needed to apply for external fellowship support had been reached. As a result, we see a modest decrease in the numbers of new fellowships in 2011–2012 when compared with the 2009–2010 and 2010–2011 years.

**Figure 5.** New F-series fellowships awarded. In 2005, 10 new F-series fellowships were awarded to graduate students or postdoctoral fellows in the BioMed division at Brown University. This value is set as a baseline of 0, and new fellowships awarded in subsequent years are incremental to this baseline. The 2008–2009 year corresponds to the start of the Brown IMSD program and its practices throughout the BioMed division.

**Faculty Engagement Driving Changes in Institutional Culture**

Brown IMSD assists graduate programs in identifying UR students as prospective trainees in training programs in which they are greatly underrepresented. Incentives provided in the form of increased program resources encourage graduate programs to identify and recruit promising UR students as graduate students. A large number of graduate faculty have become involved in IMSD activities, including leadership of training modules, service on the IMSD advisory board, participation in institutional partnership activities, and mentoring of IMSD trainees (Table 4). The success of partnership-building relationships is driven by the willingness of faculty to take active roles in sustaining these partnerships. While the activation energy for engagement can be high for some faculty, those involved recognize the value and benefits in supporting these relationships and building connections with prospective graduate students. Engagement of UR students by non-UR faculty allow the faculty to better understand the basis of assumptions many of these students make with regard to earning PhD degrees and the perceived value of the degree. Non-UR faculty also gain an appreciation that student motivation to engage in research and enter research careers may be dampened by family or cultural demands and perceptions of limited career options in the sciences. Training modules, which were initially viewed as extracurricular distractions that take away from important graduate work, have now received full “buy-in” from faculty, who recognize the merits of training modules as early intervention tools and encourage enrollment of both UR and non-UR students. An increasing number of faculty serve as module leaders, as they recognize that their involvement helps students develop critical-thinking abilities and analytical and writing skills.

IMSD also leverages its position to secure institutional support for faculty to attend scientific meetings at which large numbers of UR students present their research. This is used as a mechanism to identify and cultivate applicants for graduate admissions. Interest in fields such as ecology, in which UR students are very much underrepresented (Holland et al., 1992), can be stimulated when science is shown to be integrative and interdisciplinary. Brown IMSD supports a seminar series to invite speakers who engage in interdisciplinary and integrative scientific research. These speakers are cosponsored by graduate programs and draw speakers and audiences from diverse fields. Using our seminar series to illustrate the interdisciplinary nature of science has helped to stimulate greater interest among UR students in pursuing graduate training in areas in which they have been historically absent.

Increasingly, Brown IMSD works with graduate-training faculty to further promote the academic development of IMSD trainees to help them transition from IMSD support to other support mechanisms. This work involves identifying and preparing trainees for unique training experiences offered by T32 institutional training awards. This collaborative work has been valuable in increasing diversity among T32 trainees and has served as a model for a number of training faculty as they establish new training foci to be supported by institutional training awards. Figure 6 summarizes

<table>
<thead>
<tr>
<th>Category/role</th>
<th>Participant number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skill-based module leaders</td>
<td>23</td>
</tr>
<tr>
<td>IMSD advisory board member</td>
<td>18</td>
</tr>
<tr>
<td>IMSD mentors/trainers</td>
<td>28</td>
</tr>
<tr>
<td>Graduate program diversity liaison/contact</td>
<td>4</td>
</tr>
<tr>
<td>Outreach faculty (attends SACNAS and ABRCMS)</td>
<td>4</td>
</tr>
<tr>
<td>Partner institution–hosted visitor/guest</td>
<td>3</td>
</tr>
</tbody>
</table>

*Brown University faculty members assumed numerous leadership roles in the IMSD era related to increasing diversity in graduate education and training. The table shows the specific roles assumed and numbers of faculty involved over the life of the program. These roles and activities did not exist prior to the Brown IMSD program.

SACNAS: Society for Advancement of Chicanos and Native Americans in Science; ABRCMS: Annual Biomedical Research Conference for Minority Students.
Addressing the Challenge of Diversity

Figure 6. Number of new T32 training grants awarded above the 2005–2006 baseline. In 2005, 11 NIH T32 and T35 training grants were available to the BioMed division at Brown University. The figure shows new incremental training grants above these 11, as shown for 2006–2007, when two new training grants were awarded to the division.

In summary, we have approached our goal of increased graduate program diversity and student success across the BioMed division by systematically applying practices that were utilized by one highly diverse graduate program. Early outcome data are encouraging and suggest that these practices contribute to increased diversity across biological and public health graduate programs. In Figure 7, we incorporate the three major IMSD practices described in this report—strategic institutional partnerships, personalized student support, and faculty engagement—to form a working model for URM retention and success, with the hope that this model may serve as a basis for further research and efforts in diversifying the graduate ranks. Because each practice is relatable to the other two and can impact its success, it is recommended that all be developed in concert to achieve maximal results.

DISCUSSION

This report shows that it is possible to achieve a level of diversity in the sciences that approaches the general level of diversity of the U.S. population. We document program practices designed to improve the success of UR students in the graduate training environment and provide short-term outcome data. The early outcomes of our program show a positive correlation with URM retention and student achievements. Although the statistical significance of some practices may appear small as we report this early data assessing the impact of the IMSD program, these practices are of great importance.

Figure 7. Working model for increasing retention and success in PhD training. The model highlights specific practices (ovals) that are key components of the Brown IMSD program and their relationship to the three broad interventions and associated goals (shaded boxes) investigated in this study. Double-headed arrows between some practices indicate reciprocal effects. The vertical box for “Form Institutional Partnerships” conveys the cross-cutting nature and interplay of practices directed toward this intervention.
practical significance with regard to diversity and in changing the institutional climate and culture at Brown. Our institutional partnerships have aided URM student recruitment via both student–faculty and faculty–faculty interaction. Student support, provided through skill-based training modules and regular monitoring of academic progress through advisement and mentoring, has aided student transition from the undergraduate level to the graduate level and success in the graduate ranks. Underrepresented racial and ethnic minority students have long been documented to be under-prepared for graduate training (Summers and Hrabowski, 2006). One of the root causes for this continues to be their lack of access to resources that are often more readily available to their majority counterparts (Aud et al., 2010). However, gaps in preparation for graduate study are not unique to URM students. Both UR and non-UR students have utilized our training modules to fill these gaps. Faculty involvement in Brown IMSD, especially in leading skill-based modules and interacting with students and colleagues at partner institutions has led to shared values in diversity across our nine PhD programs. In turn, our graduate programs are now all succeeding in attracting and retaining high-quality URM students.

While we do not suggest that the strategies described here are unique to Brown, these practices, in combination with senior administration support, have been successful in advancing diversity and student success in our training environment. The practices described here are generalizable and can be expected to lead to similar outcomes when applied elsewhere. As a result, we look forward to seeing measurable advances in the representation of racial, ethnic, and other disadvantaged individuals in the scientific workforce.

ACKNOWLEDGMENTS
This work is supported by 1R25GM083270-01 and 1R25GM083270-01S1 from the National Institute of General Medical Science to N.L.T. and A.G.C. We thank Tracey Cronin, Michele Curry, and Karen Ball for data tracking and analysis and Dr. Elizabeth Harrington for access to divisional information. We also thank Dr. Xi Rossi Luo and Allison Delong for statistical analysis. We further acknowledge the commitment of our institutional partner leaders Drs. Gerry McNeil, (York/City University of New York), Chris Bazinet (St. John’s University, New York), and Mary Smith (North Carolina A&T State University).

REFERENCES
Scientific workforce diversity is critical to ensuring the realization of our national research goals and minority-serving institutions play a vital role in preparing undergraduate students for science careers. This paper summarizes the outcomes of supporting career training and research practices by faculty from teaching-intensive, minority-serving institutions. Support of these faculty members is predicted to lead to: 1) increases in the numbers of refereed publications, 2) increases in federal grant funding, and 3) a positive impact on professional activities and curricular practices at their home institutions that support student training. The results presented show increased productivity is evident as early as 1 yr following completion of the program, with participants being more independently productive than their matched peers in key areas that serve as measures of academic success. These outcomes are consistent with the goals of the Visiting Professorship Program to enhance scientific practices impacting undergraduate student training. Furthermore, the outcomes demonstrate the benefits of training support for research activities at minority-serving institutions that can lead to increased engagement of students from diverse backgrounds. The practices and results presented demonstrate a successful generalizable approach for stimulating junior faculty development and can serve as a basis for long-term faculty career development strategies that support scientific workforce diversity.

INTRODUCTION

Race, ethnic status, and underprivileged backgrounds shape social identities, which in turn influence educational performance and academic success (Steele, 1997; Spencer et al., 1999; National Academy of Sciences [NAS], 2011). The low numbers of individuals from these backgrounds, especially racial and ethnic minorities, who pursue careers in science, technology, engineering, and mathematics (STEM) is often cited as one of the factors likely to impede the United States’ future scientific progress (NAS, 2011). While the pool of prospective scientists among these backgrounds continues to grow, few end up pursuing science careers (Summers and Hrabowski, 2011).
suggesting that barriers to opportunities and participation continue to limit success.

Minority-serving institutions (MSIs), which include historically black colleges and universities (HBCUs), Hispanic-serving institutions (HSIs), and tribal colleges and universities (TCUs), are defined by enrollments of 25% or more underrepresented minority (URM) students (U.S. Department of Education, 2008). As of 2004, 58, 63.3, 52.9, and 38.6%, respectively, of black, Hispanic, Asian, and Native American students attended MSIs, placing these institutions at the forefront of educating and training of U.S. racial and ethnic minorities (U.S. Department of Education, 2008). The overall fraction of U.S. minority students at MSIs has also increased from 13.5% in 1984 to 32% in 2004 (U.S. Department of Education, 2008).

Faculty members at MSIs have greater and more regular access to URM students, making them important participants in the national effort to address the issue of underrepresentation in STEM careers. Because the impact of the MSIs is exceptional, strengthening the professional development of faculty is critical for the development of URM students. For example, HBCUs award 50% of the degrees in mathematics and 40% of the degrees in physics held by African Americans (Gasman, 2008, 2009; Burrelli and Rapoport, 2008). Moreover, although only five of the top 50 producers of African-American baccalaureates who go on to science and engineering PhDs are HBCUs, these five account for 25% of future black PhDs (Gasman, 2008, 2009; Burrelli and Rapoport, 2008). Trainees from minority communities who pursue careers in the healthcare fields go on to serve as care providers in minority communities (Murray-García et al., 2001; Komaromy et al., 1996). Similarly, a larger percentage of black faculty members who hold science, engineering, and health doctorates—many who often began their own careers in minority communities—often pursue careers in the scientific fields at MSIs and in minority communities (National Science Foundation [NSF], 2011). As a result, they are directly involved in addressing the problems and challenges faced by their communities while contributing to the education and diversification of our national workforce.

The Minorities Affairs Committee (MAC; www.ascb.org) is a standing committee of the American Society for Cell Biology (ASCB) that began its work in 1980. Its membership includes faculty from diverse ethnicities and scientific backgrounds and from both MSI and non-MSI teaching-intensive and research-intensive institutions. Past and current committee members are also prominent scientists, educators, mentors, and role models. In addition to supporting the mission and programs of ASCB, MAC’s unique mission is to increase diversity in cell biology and address issues affecting underrepresented racial and ethnic minorities in the sciences. This is achieved by supporting educational training and the participation of minorities at all leadership levels in the society, and by working to improve the professional development of faculty members, especially those at MSIs. This is being achieved through programs such as the Linkage Fellows Program, the Junior Faculty and Postdoctoral Fellows Career Development Workshop, and the Visiting Professorship (VP) Program.

Faculty members establish the conceptual frameworks in which students learn. The development of students who are both reflective and inquisitive in the classroom is dependent on access to educators who are themselves also reflective and inquisitive and who possess a good understanding of the practice of science inquiry (Minstrell and van Zee, 2000; Drayton and Falk, 2006). Further, the growing appreciation for the impact of inquiry-based learning, including research-based laboratory courses (Garde-Hansen and Calvert, 2007; President’s Council of Advisors on Science and Technology, 2012), means that faculty research training and experience will become an increasingly important prerequisite for effective biology education (Feldman, 1987). In response to these needs, the VP Program was designed in part to provide research opportunities to faculty members who spend the majority of their time in the classroom, to enable them to become better able to improve curricular science content and pedagogy.

The MAC developed the VP Program to meet the professional scientific needs of faculty members at MSIs by providing research training to enhance scholarly practices with the goal of strengthening educational and research activities at their home institutions. In assessing the effectiveness of the VP Program, we asked the following questions: 1) Does the program enhance MSI faculty scholarly practices? 2) Does the program impact the teaching practices of the faculty participants? The present paper reports on the outcomes for 32 Visiting Professors and compares their progress with that of 129 matched MSI faculty peers.

METHODS

The VP Training Program

The VP Program was established in 1997 with the support of a Minority Access to Research Careers grant from the National Institute of General Medical Sciences. The program helps to strengthen research infrastructure and teaching practices at MSIs where large numbers of underrepresented racial and ethnic minority students receive their undergraduate degrees and research training. It also provides faculty members with professional development opportunities not readily available at their home institutions by enabling them to engage in research activities that increase their ability to publish scientific works, present their findings at regional and national scientific meetings, and develop extramurally funded research programs. The program also supports the professional development and enhances the career trajectories of the participants by supporting teaching and curricular practices and by encouraging the faculty members to expand their professional networks. Participants receive 8- to 10-wk training internships in the laboratories of host scientists at research-intensive institutions. Faculty sponsors are often ASCB members, and all maintain active research programs. These faculty sponsors are accomplished scholars who have both laboratory and classroom training expertise. Their research interests also align well with the goal of the program and mission of MAC and the ASCB. In any year, a faculty sponsor may host only one VP Program participant. Sponsoring faculty and Visiting Professors are not required to be members of ASCB to participate in the VP Program.

Selection of VP Program Participants

Calls for applications and nominations to the VP Program are made to faculty at research-intensive institutions and
MSIs. Application calls are also made through published notices distributed at scientific meetings, the ASCB website and newsletter, email announcements, and advertisements in specialized scientific journals. Applicants are drawn from a national pool, and applications are reviewed by the MAC VP selection committee comprising several members of the MAC. Participants selected for the program are those who hold academic appointments at MSIs, especially teaching institutions, throughout the United States and its territories. Of the 32 program participants described in this paper, there are 19 African Americans, four Asians, three Caucasians, and six Hispanics. Nineteen are female. At the time of their participation in the program, all were employed as full-time regular faculty with academic duties that included classroom teaching of undergraduate students. Visiting Professors carried out cell biology–based laboratory research with faculty mentors in the summers, during which time they received a stipend and a modest allowance to support the research activities.

All VP applicants are required to outline a research plan as well as the nonmaterial support and mentoring they will receive from sponsors. The successful outcome of this plan is expected to be the launching of independent research programs by the Visiting Professors, which will allow them to engage students at their home institutions in the classroom and in teaching and research labs. Consideration is given to both short-term and long-term goals and objectives of the applicants, their skill sets, and the value of the scientific outcomes of the work to be completed. Consideration is also given to the infrastructures and capacities at the applicants’ home institutions, because those resources may not be equivalent to those at the sponsors’ institutions. In particular, consideration of the feasibility of continuing some aspect of the work at the Visiting Professor’s home institution is evaluated, including limitations related to instrument availability and experimental model system costs and availability. Faculty members may participate in the VP Program more than once, but participation thus far has been limited to a maximum of two summer training experiences, with preference given to junior untenured scientists in tenure-track positions or en route to tenure-track positions.

Assessment of Participants and Program Outcomes

From 1997 to 2012, 43 unique VP Program participants completed 60 VP experiences (Figure 1). The present evaluation is limited to participants who were at least 1 yr beyond their first VP experience. This represents a group of 32 unique scientists participating in 49 VP experiences, with 17 of 32 participating in the program twice through 2011. Because the number of applicants to the program who were awarded but declined VP support is small, it was not possible to use this group for accurate comparisons with program participants. Accordingly, matched faculty peers at the same academic ranks and at the same home institutions and in the same departments as VP Program participants were selected as the control group. The period 5 yr prior to participation in the VP Program and the period after participation were compared for publications and grants authored by Visiting Professors and by a group of 129 matched MSI faculty peers.

Self-reported and public online data were collected for the VP Program participants during the period 1997–2011. Database searches were performed using PubMed (www.ncbi.nlm.nih.gov/pubmed?db=pubmed) to access publications from MEDLINE available through 2012. Information on federal grant funding was collected from publicly accessible databases, including the National Institutes of Health (NIH) RePORTER database (http://projectreporter.nih.gov/reporter.cfm), the NSF Award Search database (www.nsf.gov/awardsearch) and the U.S. Department of Agriculture (USDA) National Institute of Food and Agriculture (NIFA) database (http://cris.csrees.usda.gov/menu.html). For funding analysis, unique grants held by Visiting Professors and controls, as either principal investigator or co–principal investigator, were included. A similar comparison was performed after Visiting Professors completed their training and includes analyses of funding data available through 2012. The number of peer-reviewed publications in the 5 yr prior to their participation in the VP Program was compared with publications during the same period of their matched MSI faculty peers. Similar comparative analyses of publications “post-VP training” were also carried out. The 5-yr period of funding analysis and publication analysis that preceded the VP training experience of each VP Program participant is referred to as the “pre-VP era.” The period after which participants have completed their VP training is referred to as the “post-VP era.” This period has been as short as 2 yr for participants who completed their VP training in 2009, and as long as 14 yr for participants completing training in the 1997–2000 period. Grants awarded to matched MSI faculty peers or to participants in the year that an individual began the VP Program were assigned to the pre-VP era. Similarly, manuscripts published in the year that faculty members became participants of the VP Program were assigned to the pre-VP era.

Data on curricular and other ancillary activities reported here represent participant self-reported data. They were collected through confidential online surveys and interviews at the conclusion of training that provided Visiting Professors...
the opportunity discuss experiences, work products, future research, and teaching plans and to provide feedback on the strengths and weaknesses of the program.

**Overview of Data Analysis and Statistical Method**

Collected data were used to evaluate accomplishments of program participants in areas that define professional success and tendency toward success in the sciences. The evaluation criteria include a measure of the average numbers of publications and average numbers and sizes of federal grants. Other posttraining activities, including curricular activities, are also reported, because these also serve as measures of the impact of VP Program training on faculty professional development and activities that impact student training and learning.

Mixed-effects regression modeling was used to evaluate whether the number of publications, number of funded grants, and grant-funding amounts following the VP experience differed for Visiting Professors compared with their MSI-matched, near-peer colleagues. The parameter estimates of interest were the fixed-effect estimate for being a Visiting Professor. Each model was adjusted for the measurement in the 5-yr period prior to program enrollment (fixed effect), and individuals were nested within institutions (random effect). The number of publications and grants were assumed to follow a Poisson distribution, and the total grant amounts were assumed to be normally distributed after log (x + 1) transformation.

The period of data analysis following VP training for Visiting Professors and their matched peers were identical. This period was as short as 2 yr for participants of the program in 2009–2010 and 2010–2011 in the 2009–2013 cohort and as long as 14 yr for the 1997–2000 cohort.

**RESULTS**

**Characteristics of Program Participants as a Function of Time**

Since the VP Program’s inception in 1997, the number of participants has grown steadily. With the exception of one minority faculty participant from a majority-serving, research-intensive institution, all participants came from MSIs. For the 1997–2000 period, six trainees participated in the program, which compares with 32 participants for the 2009–2011 period. Although the participation data for the 2009–2013 period are incomplete, the number of participants through 2011 was twice that of the prior 4-yr period. Because the number of applicants to the program has always outpaced available slots (unpublished data), individuals invited to participate in the program are selected through a competitive process.

None of the participants in the program to date has reported on his or her disability status, and therefore only data on racial and ethnic status are available (Figure 2). African Americans and Hispanics account for the majority of participants and applicants to the program (unpublished data). While there has been a steady increase in applications and participants from HBCUs, there has been a less consistent rise in the number of applications from faculty members at HSIs. Asian Americans accounted for a small fraction of applicants to the program and were only represented in the program in the 2001–2004 and 2009–2013 periods. Although one Native American faculty scientist served as a co-mentor and co-host, no Native American faculty members have participated in the program as a sponsor or as a Visiting Professor. The VP Program supports faculty at MSIs regardless of race and ethnic background, and there has been steady increase in the number of participating non-URM faculty members in the program in the last 5–7 yr. From 2009 to 2011, non-URM faculty members accounted for 17% of all participants. This compares with 7% for the 2005–2008 period and 0% for both the 1997–2000 and the 2001–2004 periods. Women have been well represented in the VP Program, accounting for 60% of all participants. For each 4-yr period of analysis, female participants outnumbered males.

**Participant Publications**

The 32 program participants reported publishing 91 manuscripts or book chapters and cite 39 of their published manuscripts as coming after their VP training experiences. The majority of these publications were generated within 3–5 yr of completing their training. A number of the publications self-reported by VP Program participants, however, included publications with postdoctoral advisors and have attribution to their postdoctoral training institutions. Although useful in measuring productivity, these works do not accurately reflect the independent career achievements of participants or the impact of the VP Program, because they are not all related to VP training experiences. For more accurate assessment of the impact of VP training on participant achievements, refereed publications generated during the pre-VP and post-VP eras by participants while in their current positions were retrieved from PubMed and compared with publication output by matched MSI faculty peers. These data are summarized in Table 1. In the pre-VP era, the average manuscript output per Visiting Professor was the same as the output for the MSI faculty peer controls (0.84 and 0.85). In contrast, during the post-VP era, the average number of manuscripts produced per participant increased to 1.37, versus 0.82 for the matched MSI faculty peers (p = 0.004). The change in SD (or σ) values for publications by Visiting Professors in the pre-VP era (1.90)
Table 1. Comparison of average number of publications and research grants per Visiting Professor versus matched MSI faculty peers

<table>
<thead>
<tr>
<th>Period</th>
<th>Average number of publications per matched MSI peer ((n = 129))</th>
<th>Average number of publications per VP Program participant ((n = 32))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-VP</td>
<td>0.85</td>
<td>0.84</td>
</tr>
<tr>
<td>Post-VP</td>
<td>1.37</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Period</th>
<th>Average number of new grants per matched MSI peer ((n = 129))</th>
<th>Average number of new grants per VP Program participant ((n = 32))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-VP</td>
<td>0.3</td>
<td>0.06</td>
</tr>
<tr>
<td>Post-VP</td>
<td>0.16</td>
<td>0.59</td>
</tr>
</tbody>
</table>

*aCollected data represent data available through 2012.

*bPublications. PubMed was used to access science journal publications from MEDLINE. Publications 5 yr prior to participation in the VP Program and only for faculty members at their current institution vs. post-VP publications were compared (95% CI 2.22–8.98, \(p < 0.004\)). The SD (or \(\sigma\)) for publications by matched peers in the pre-VP era and post-VP era were 1.84 and 2.10, respectively. For publications by Visiting Professors in the pre-VP era and post-VP era, these values were 1.9 and 2.37, respectively.

*cResearch grants: Using NIH, NSF, and USDA funding databases, federal grants held 5 yr prior to faculty participation in the VP Program vs. post-VP successes were compared (95% CI 1.18–2.44, \(p = 0.004\)). The SD (or \(\sigma\)) for publications by matched peers in the pre-VP era and post-VP era were 0.78 and 0.41, respectively. For publications by Visiting Professors in the pre-VP era and post-VP era, these values are calculated to 0.24 and 1.38, respectively.

versus the post-VP era (2.37) is small. Similarly the change in SD for publications by the matched MSI peers is also observed to be small, going from 1.84 in the pre-VP era to 2.10 in the post-VP era. Figure 3 shows the distribution of post-VP training publications of program participants over time. Approximately half of all publications attributed to VP Program participants were produced by participants who completed their training from 1997–2000. The 2001–2004 support period represents the period in which the lowest publication output by Visiting Professors is reported (Figure 3). This also aligns with the lowest level of publication output by Visiting Professors in the pre-VP era and the lowest number of new grants secured by Visiting Professors in the post-VP era. Although not a favorable outcome, this observation is consistent with decreasing numbers of articles produced by U.S. scientists per federal grant beginning in 2001 (Boyack and Jordan, 2011). It also coincides with the start of decline in NIH grant success rates for both Research Projects Grants (RPGs) and R01 Equivalent Awards (R01s). These values decreased from 32.1% for RPGs and 31.7% for R01s in 2001 to 25% and 24.6% in 2004, respectively, representing the largest-ever drop and rate of decrease in federal grant funding (Federation of American Societies for Experimental Biology, 2013). It was not possible to establish causal relationships between funding and publications. Overall, VP research training correlates positively with posttraining publications, and individuals who participated in the program more than once produced 80% of their manuscripts in journals of equal or higher impact factors as those of their MSI peers (unpublished data).

**Grant Funding**

The pairing of VP Program participant and host scientist is based on common research interests and expertise, and these pairings have resulted in research collaborations. More than one-third of participants report that they collaborate with or are in the process of establishing collaborations with their host sponsors, which will help to sustain the momentum of work initiated during the VP training experience. As part of these collaborations, some participants report jointly submitting grant applications with their past sponsors.

Many VP Program participants attribute securing grants that support their research and teaching practices to their VP training. Of the past 32 Visiting Professors, 13 reported grant funding for work directly related to their VP training experience. Of this group, 10 reported having received grants for work on which they serve as principal investigator or as co–principal investigator. These include federal research grant awards, professional development grants, and teaching grants. An additional three participants reported receiving grant funding in the form of state or private grants; only two of the past 32 participants reported that they did not apply for grant funding following the VP training experience. The success in securing federal grants by Visiting Professors, which is an indicator of funding success and is concordant with the federal funding mechanism used to support the VP Program, was examined (Table 1). Grant support correlated positively with VP training, and participants attribute securing grants supporting their research and teaching practices to their VP training. Forty percent of past Visiting Professors credit funding for grants on which they serve as principal or co–principal investigators to their VP training experiences. Together, the 32 faculty members who would eventually
become Visiting Professors held two federal grants (NIH, NSF, NIFA) in the pre-VP era (average of 0.06 grants per person); their number of grants increased nearly 10-fold to a total of 19 in the post-VP era (average of 0.59 grants per person). In comparison, the 129 MSI faculty peers comprising the control group held 39 federal grants (0.30 grants per person) in the pre-VP era, and decreased to a total of 19 grants in the post-VP era (average of 0.15 grants per person).

The average amount of federal grant dollars available per Visiting Professor in the pre-VP era totaled ~$17,530 compared with ~$117,180 in the post-VP era (Table 2). For the matched peers, the post-VP-era new grant total was $124,000 compared with $202,000 in the pre-VP era. The smaller average size of MSI grants, compared with what is reported for principal investigators at research-intensive institutions (NIH, 2012b), is consistent with the smaller scale of the research programs and different mission at MSIs.

### Posttraining Activities Related to Student and Professional Development

VP Program participants reported on a number of other activities and outcomes that have been influenced by their training experiences, and these are summarized in Table 3. Research training has influenced teaching practices by program participants, as 22 of the past 32 participants reported developing new courses or instituting curricular changes that integrate content from their VP experiences. Curricular improvements reported include introduction of cutting-edge methodologies such as polymerase chain reaction technology into courses, establishing independent research projects as components of courses, and use of journal articles and research proposal writing exercises. Other curricular upgrades and revisions included addition of online components to teaching practices.

As indicated in Table 3, 31 of 32 VP Program participants report that they mentor undergraduate students. Nineteen participants report that they have assumed leadership roles at their home institutions and in the larger scientific community. These roles range from becoming research advisors to graduate programs, to joining the ASCB Minorities Affairs Committee as regular and ad hoc members, to organizing scientific workshops at regional scientific meetings. Participants also report becoming more active in scholarly pursuits relating to their professional development. In posttraining surveys, 22 participants report attending scientific meetings and more than half of these report attending the annual ASCB meeting at least once every 2 yr. Twenty-one report being members of professional societies, with 17 being active members of ASCB.

Tenure and promotion among VP Program participants over the life of the program was examined, and the results are summarized in Figure 4. Sixteen of the 32 faculty participants of the program through 2011 were tenure-eligible faculty members, and 14 of the 16 were awarded tenure following their VP experience. The majority of these advanced from the rank of assistant professor to associate professor and reported being tenured within 5 yr after their VP training.

### Table 2. Comparison of average size of research grants (in 1000s of US dollars) per Visiting Professor versus faculty peers

<table>
<thead>
<tr>
<th>Period</th>
<th>Average grant size per matched MSI peer</th>
<th>Average grant size per VP Program participant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-VP</td>
<td>202</td>
<td>17.53</td>
</tr>
<tr>
<td>Post-VP</td>
<td>124</td>
<td>117.18</td>
</tr>
</tbody>
</table>

*Using NIH, NSF, and USDA funding databases, grants held 5 yr prior to faculty participation in the VP Program vs. post-VP successes were compared. Comparisons were made between Visiting Professors who were matched against their peers in the same department in the same home institution. In total, 32 Visiting Professors were compared with 129 peer faculty. Research grant funding directly and indirectly related to work completed by Visiting Professors. Grants identified were only new grants by faculty since completing the VP Program and represent grants on which participants were either principal investigators or co–principal investigators. Collected data include only data available through 2012.

### Table 3. Ancillary activities of Visiting Professors following training

<table>
<thead>
<tr>
<th>Activity</th>
<th>Number of participants (32 total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research collaborations with host scientist</td>
<td>13</td>
</tr>
<tr>
<td>New course development or curricular improvements</td>
<td>22</td>
</tr>
<tr>
<td>Student research training and mentoring</td>
<td>31</td>
</tr>
<tr>
<td>New leadership roles</td>
<td>15</td>
</tr>
<tr>
<td>Attendance at professional scientific meeting</td>
<td>22</td>
</tr>
<tr>
<td>Professional society membership</td>
<td>21</td>
</tr>
</tbody>
</table>

*Participants of the VP Program reported on activities at their home institutions and in the larger scientific community related to training and professional development upon the completion of the program. Thirty-two participants were queried. Activities engaged in are listed here and the number of participants engaging in each activity is presented.

Figure 4. Academic promotion of past participants of the VP Program. Tenure-eligible faculty members who participated in the VP Program self-reported on their tenure status following completion of the program; self-reported are summarized here. The period listed corresponds to the training period of each participant.
Grant-funding successes by Visiting Professors correlated positively with VP training (Tables 1 and 2). Although MSI faculty peers were also successful in securing grants in both the pre-VP and post-VP era, there was an ~50% decline in new grants received by this peer group in the post-VP era. The larger number of grants held by Visiting Professors in the post-VP era is likely a result of VP training, as many of the Visiting Professors report writing collaborative grants with their sponsors and often return to their home institutions energized to sustain their research practices. Many also are likely to have spent more time writing grants during the post-VP era than their peers, as they did not hold the level of funding held by their matched peers. The increase in the SD (from 0.24 in the pre-VP era to 1.38 in the post-VP era) for numbers of new grants secured by Visiting Professors indicates significant variation in productivity among the participants; this might be in part a discipline-specific effect, wherein the amount of funding available and opportunities for funding may not be equivalent in the fields in which Visiting Professors work.

There is a positive correlation between participation in the VP Program and subsequent grant funding. While the average grant size per Visiting Professor increased in the post-VP era, the average grant size for their matched peers decreased from the pre-VP era to the post-VP era. This change is not statistically significant. One peer scientist held an extremely large grant, and thus the mean difference between the two groups was not significant, despite Visiting Professors having a greater rate of funding. Because of the limitations in the survey data, we do not know the frequency with which each applied for funding, nor do we have a complete view of success rates on first submissions. Overall, the federal grant-funding success rates and average sizes of grants among MSI faculty are difficult to calculate, and it is therefore difficult to place the values presented above into a larger context.

Data presented show that participation in the VP Program correlates with other positive posttraining activities and practices. The VP Program participants reported engaging in activities that support their individual professional development and the development of their student trainees. Visiting Professors have also reported that they have been able to bring their experiences into their teaching classrooms. Although the extent to which individuals with bench research experience bring direct benefit to classroom teaching remains debatable (Smeby, 1998; Marsh and Hattie, 2002; Pocklington and Tupper, 2002), their involvement in classroom biology education does support didactic and other teaching methods (Miller et al., 2008).

Finally, the majority of Visiting Professors eligible for tenure and promotion were tenured and promoted. We do not know the significance of this, due to the small number of candidates in this group and the absence of tenure and promotion information for their peers.

It is likely that the effectiveness of the VP Program is due in no small part to the program being one of a suite of opportunities offered to these scientists. Other highly structured, nonresearch opportunities offered by the MAC are among the important foundations and contributors to participant success. Seventeen of the past 32 Visiting Professors have participated in the MAC Linkage Fellows Program, which provides support for outreach activities in cell biology at participants’ home institutions. Twelve past Visiting Professors
have participated in the Junior Faculty and Postdoctoral Fellows Career Development Workshop, which provides career development training. In 2005, participation in this workshop became mandatory for all Visiting Professors. Visiting Professors are also required to present their work at the annual ASCB meeting. Together, these activities can reinforce the scientific identities of the participants by building self-efficacy and a sense of belonging to the scientific community (Estrada et al., 2011).

At the undergraduate level, there is strong evidence that authentic research experiences reinforce scientific identity and the pursuit of a career in science (e.g., Nagda et al., 1988; Hathaway et al., 2002; Russell et al., 2007; Thiry et al., 2011; Hernandez et al., 2012). Similarly, faculty development programs build the networks that support faculty academic success, and faculty lacking access to career development resources lack the foundations for professional success (Hitchcock et al., 1995; Morzinski and Fisher, 2002). The focus of the VP Program is to engage faculty at teaching-intensive institutions in research activity, thus re-establishing the framework for appropriate career development. Strengthening this framework—comprising skills such as motivation, persistence, mentoring—reinforces the scientific identity of the participating Visiting Professor. More important is the potential lasting and positive effect on the Visiting Professor’s students.

The information gathered through the VP Program represents the initial step in longitudinal and retrospective studies that will be useful in defining challenges to individual scientific career development, especially for those from underrepresented groups and faculty at MSIs. Participation in the VP Program has been shown to lay the foundations for faculty members to achieve individual long-term career goals by increasing their engagement in research activities, which impact their teaching practices. Figure 5 presents a model that highlights the important gains linking mentored research practice to the desired outcomes for participants of the program. This model serves as a guide to assist faculty mentors and sponsors to strengthen the development and success of junior faculty. This is especially true at MSIs, where faculty plays an important role educating and preparing URM students for careers in STEM fields.

The disparity in research funding between majority and minority scientists (Ginther et al., 2011) was the focus of a recent report (NIH, 2012a). In response to this report, the NIH has recently announced new initiatives, including a renewed emphasis on underresourced undergraduate institutions on the premise that these institutions can produce more students prepared for research careers (http://commonfund.nih.gov/diversity/initiatives.aspx). A key element in the improved preparation of undergraduate students is the ability of the faculty to teach through inquiry and to provide students opportunities to engage in authentic research (Hernandez et al., 2012). The VP Program offers important insights that can inform the design of effective strategies to enhance the development of MSI faculty. Among the elements that are key to the success of the participants, we present three here: 1) a modest annual financial investment of ~$6000 per participant; 2) the commitment of host scientists; and 3) the willingness of the Visiting Professor’s home institution to allow the faculty member to explore new ways to introduce inquiry-based learning into students’ experiences. The patient, often person-by-person investment in faculty development promises to ultimately translate into educational gains that benefit the larger scientific community in terms of scientific productivity and workforce development.

ACKNOWLEDGMENTS

This work is supported by T36GM008622 from the National Institute of General Medical Science to the ASCB and R25GM083270 to A.G.C. Human subject data collection and the writing of the manuscript were supported by grant R25GM083270 to A.G.C. The authors thank ASCB Senior Manager Deborah McCall for administrative and technical support and information gathering. We also thank Joy Quill of
C. J. Quill & Associates, Inc., for data tracking and analysis and Alison Delong of the Brown University Center for Statistical Sciences for assistance with statistical analysis. We further acknowledge the commitment and contributions of current and former members of the MAC and the leadership of the ASCB.

REFERENCES


CBE—Life Sciences Education (LSE) is an online, quarterly journal owned and published by the American Society for Cell Biology (ASCB). The journal publishes original, previously unpublished, peer-reviewed articles on life sciences education research and evidence-based practice at the K–12, undergraduate, and graduate levels. The editorial board of LSE is particularly interested in students’ conceptions and cognition related to biology and scientific inquiry and students’ engagement in the study of life sciences, as well as the efficacy of particular approaches to teaching in the life sciences. In addition, the ASCB believes that biology learning encompasses diverse fields, including math, chemistry, physics, engineering, and computer science, as well as the interdisciplinary intersections of biology with these fields. One goal of the journal is to encourage teachers and instructors to view teaching and learning the way scientists view their research, as an intellectual undertaking that is informed by systematic collection, analysis, and interpretation of data related to student learning. Target audiences include those involved in education in K–12 schools, two-year colleges, four-year colleges, science centers and museums, universities, and professional schools, including graduate students and postdoctoral researchers. LSE is published online four times a year: March (Spring issue), June (Summer issue), September (Fall issue), and December (Winter issue). LSE also prints a highlights issue each year in December, featuring contributions selected from the four online issues. Submissions are accepted at any time. Articles are assigned to particular issues by the editors. To be included in an issue, manuscripts must be accepted in final form at least two months prior to the publication date.

All published articles are available freely online without subscription. In addition, published articles are indexed in PubMed and available through PubMed Central. A goal of LSE is to stimulate dialogue. LSE invites readers to send comments that will be considered for publication in the Letters to the Editor column.

**Determining the Suitability of a Manuscript for LSE**

**Articles.** LSE publishes articles that focus on questions of science education research and practice. Articles can either break new ground in understanding teaching and learning or describe the implementation and evaluation of educational innovations. Articles must include collection, analysis, and interpretation of data, which can be quantitative or qualitative in nature. In addition, LSE articles should: 1) address a clear educational problem or education research question, 2) demonstrate clear alignment among the problem or question being addressed, the design of the study or educational innovation, the claims being made, and the evidence used to support those claims, 3) describe how results are applicable or transferable to other settings, 4) be relevant to a defined audience of educators, and 5) make reference to related educational literature. When possible, articles should include any newly described instructional resources and assessment instruments so that they are available for evaluation by reviewers and use by readers. The online nature of the journal facilitates the inclusion of instructional materials such as syllabi, assignments, rubrics, laboratory protocols, or professional development guidelines.

Articles detailing educational innovations must include assessment appropriate for the innovation as well as the assessment instruments, protocols, rubrics, and methods, or references to these components of the work. LSE is interested in educational innovations that have not been described extensively elsewhere, although articles about the outcomes of adopting or adapting particular approaches to teaching and learning with diverse learners and in diverse learning contexts are welcomed. Carefully constructed descriptions of relevant literature are important for demonstrating how a particular innovation is unique compared with previously published work. Instructors interested in publishing their educational innovations in LSE should give careful thought to how they will assess student learning. Articles that lack adequate assessment, assessment instruments, descriptions of assessment methods, or references to published assessment instruments or methods will be returned to authors without review.

Because the concept of assessment may be unfamiliar to many instructors, some suggestions for assessment are presented here. Answering three questions can help guide the process of assessment: 1) What are your instructional goals? 2) What should your students know or be able to do if you met your goals? 3) How can you measure or otherwise document whether your students know or are able to do what you intend? Documentation of intended outcomes can be accomplished through systematic analysis of data collected through diverse approaches, such as pretest/posttest, interviews, focus groups, surveys, or performance on coursework, including exams, papers, or lab reports. Authors should present their innovations in the same way that life scientists present their research: claims regarding efficacy should be supported by evidence.

Answering the following questions can help you determine whether your manuscript is ready for consideration by LSE: What is the background or motivation for your project? Does it meet a previously unmet need or provide a test of some educational theory? What was your goal in developing it? How do you know if you have succeeded? Can you show success to readers? Are the outcomes and supporting evidence sufficient to convince others to adopt your materials, approach, or project for their own use? Are the outcomes sufficiently novel to warrant publication?

**Essays.** LSE publishes essays, which describe approaches to educational challenges, including assessment methods, student engagement, curriculum innovations, K–20 continuum, and other topics. Essays are framed by personal experience and provide specific examples, but describe a problem or approach of general interest and may be synthetic across the work of many individuals. Appropriate foci for essays include reviews of current practices, policies, or research that have implications for biology teaching and learning, or personal perspectives on issues that are provocative or would otherwise be of widespread interest. The problem or approach should be presented within a scholarly context, citing references and resources that address the topic. If claims are to be made, there should be evidence from the literature or the authors’ own work. Although it is not a requirement, essays can include ideas for assessment or future research as appropriate.

**Meeting Reports.** The journal publishes occasional reports of meetings related to education in the life sciences. Meeting reports received through the normal online submission process will be judged on the basis of their relevance and interest for the broad readership of the journal.

**Letters to the Editor.** LSE welcomes letters as a means for initiating discussion of issues of interest to readers. LSE requires
that letters are directly responsive to an article published in the journal or that they bring a new or under-recognized issue to the attention of readers in a way that is informed by the inclusion of relevant data or references to the literature. Letters received through the normal online submission process will be reviewed by the Editor-in-Chief and published at his discretion. Appropriate topics for letters include commentary on education-related articles in this or other journals as well as opinions on more general topics of concern to the readership.

Features. Articles listed under the heading of “Features” are by invitation only. Authors interested in contributing to a feature should contact the editor.

Announcements. LSE accepts noncommercial announcements of meetings, workshops, and conferences and of funding opportunities and fellowships open to all.

Peer-Review Process

All submitted manuscripts and educational materials are subject to peer review. After a manuscript has been submitted to LSE, the Editor-in-Chief selects an editorial board member to guide the paper through the review process. Editorial board members select two reviewers to submit written evaluations. The board member will assess the peer reviews and determine whether the submission will be accepted as is, accepted with suggested revisions, temporarily rejected with suggestions for improvements before resubmission, or rejected with reasons explaining this decision. The corresponding author can usually expect an initial response within four weeks.

License and Publishing Agreement

Authors are required to sign a License and Publishing Agreement when a manuscript is accepted for publication. Under the terms of that agreement, authors retain copyright but grant the ASCB a perpetual license to publish the manuscript. Authors also grant to the general public the nonexclusive right to copy, distribute, or display the manuscript subject to the terms of the Creative Commons–Noncommercial–Share Alike 3.0 Unported license (http://creativecommons.org/licenses/by-nc-sa/3.0).

ASCB Policy on Research Misconduct by Authors

By submitting a paper to LSE, an author acknowledges that he or she is subject to the ASCB Policy on Research Misconduct by Authors. The policy is posted at www.ascb.org/files/research_misconduct.pdf.

Institutional Review Board Approval

Manuscripts reporting on studies involving human subjects must include explicit assurance that the research reported was approved by a local Institutional Review Board, including an Institutional Review Board number, unless the research is exempt from such review according to U.S. Department of Education guidelines. Prospective authors are advised that permission must be obtained in advance.

Guidelines for Preparing Articles, Essays, and Features

General Instructions. Submitted papers must be original (i.e., not published or submitted for publication elsewhere). Previously published material will not be considered. Publication of a short abstract does not constitute prior publication, nor does presentation of data at a scientific meeting or in a webcast of such a meeting. The posting of a manuscript on an author’s personal website or in an online institutional repository prior to submission of the manuscript to LSE is not considered prior publication. Publication of a paper in the proceedings of a scientific meeting generally does constitute prior publication.

The text should be written in clear, concise, and grammatical English. Manuscripts ordinarily begin with an overview of how the work presented is relevant to the classroom, laboratory, or curriculum and what student outcomes are expected. Whenever possible, incorporate materials by citing relevant publications, without repeating already published works. Manuscript files must be submitted in either .doc or .rtf format.

Tables and illustrations should convey information effectively. Graphs and figures should be provided digitally as separate TIF or EPS files. Authors are encouraged to take advantage of the online nature of LSE. Video, audio, databases, images, animations, molecular structures, and other electronic resources may be linked as supplemental material for further consideration by readers.

International authors may wish to consider using an editorial service, such as ScienceDocs (www.sciencedocs.com), The Medical Editor (www.themedicaleditor.com), American Journal Experts (www.journalexperts.com), Editage (www.editage.com), Bioscience Writers (www.biosciencewriters.com), or Squirrel Scribe (www.squirrelscribe.com). LSE does not endorse any particular service, and cannot be held responsible for the services they provide.

Length Guidelines. The following manuscript submission lengths are intended to aid authors in preparing their manuscripts; however, submissions outside these ranges will be considered.

Articles: 30,000–60,000 characters (with spaces), or 5–10 journal “pages”; typically do not exceed 20 journal pages, or 120,000 characters.

Essays: 12,000–30,000 characters (with spaces), or 2–5 journal “pages”; typically do not exceed 10 journal pages, or 60,000 characters.

Features: 6,000–12,000 characters (with spaces), or 1–2 journal “pages.”

Letters: 3,000–6,000 characters (with spaces), or up to half a journal “page.”

Cover Letter. Authors should submit a cover letter from the corresponding author stating that the work is being submitted exclusively to LSE and indicating why it is appropriate for the journal. If there is a connection between an author and a commercial product being used or reported, full disclosure is required in the cover letter and appropriate statements should be included in the manuscript. (See “Title Page” below.) Authors are invited to suggest reviewers (please include institution and email address).

Title Page. Page 1 should include the title of the manuscript, the type of manuscript being submitted (e.g., article, essay, feature, letter to the editor, response), the number of characters in the manuscript, a shortened running title (not to exceed 42 characters and spaces), and the names and
affiliations (including department, institution, city, state, and zip code) of all authors in the order in which they should appear. List the corresponding author separately with complete postal and email address and telephone and fax numbers. Keywords should also appear on page 1. Include at least five keywords selected from the text of the article. If possible, include a keyword that indicates the target student group (primary, secondary, undergraduate, graduate, or postdoc) and a second keyword that indicates which cells or organisms are studied in the course or lab exercise (e.g., bacteria, fungi, plants, yeast, Drosophila, nematodes).

If one or more of the authors of a research paper that assesses the effectiveness of a product or curriculum, readers need to be fully aware of this potential conflict of interest. Therefore, any potential conflicts of interest should be clearly stated on the title page of the manuscript. The author and the product should be identified, and a statement included that no promotion of a particular product to the exclusion of other similar products should be construed. This will be noted under the byline if the manuscript is accepted for publication.

Abstract. Page 2 should contain the abstract, which should be no more than 200 words long and should summarize the important points in the manuscript.

Manuscript Text. The text of the paper should begin on page 3. LSE follows the style guidelines of the Council of Biology Editors Style Manual. For chemical nomenclature, follow the Subject Index of Chemical Abstracts. Capitalize trade names and give manufacturer names and addresses. Do not include figures or tables within the body of the manuscript. A format of Introduction, Methods, Results, Discussion, and References is encouraged, but other formats may be more appropriate for some topics. All pages must be numbered.

Accessing Materials. Describe how to access new educational materials if the study or use of such materials is the subject of the paper. If materials are online, provide a URL to the material. Any registration requirements or agreements inherent in the use of the materials should be described. If there are no online materials, simply state “No additional materials available online.” For other new educational materials presented in the manuscript, authors should describe how readers can access the materials, what format is available (e.g., DVD, CD-ROM, PDF files, and html pages), how to request copies, and if there are any costs. LSE encourages provision of materials on a nonprofit basis, but recognizes that this is not always feasible. Please contact LSE with any questions regarding this policy.

Acknowledgments. Identify financial sources and other sources of support for the research being reported in the manuscript.

References. Place the reference list immediately following the manuscript text (beginning on a new page). Consult the most recent issue of LSE for reference citation formats. Cite references in the text alphabetically by name and date (e.g., Jones and Smith, 1987; Smith et al., 1988). Only published articles or manuscripts accepted for publication can be listed in the Reference section. References should contain complete titles and inclusive page numbers and should be listed in alphabetical order. Abbreviate the names of journals as in PubMed. The abbreviations can be found at LinkOut Journal Lists.
Guidelines for Preparing Electronic Resources

It is possible for authors to submit for peer review electronic works including, but not limited to, animations, Chime tutorials, movies, interactive websites that may include quizzes, images (electron micrographs, photomicrographs, etc.), Java Applets, searchable databases, etc. Articles that describe new educational uses of existing resources (e.g., Expression Connection, FlyBase, Database of Interacting Proteins, etc.) are also of interest. Manuscripts of this type should provide detailed instructions for use of the resource by the target audience.

For works using someone else’s electronic resource (such as a database), a letter from the creator or curator of the resource indicating their willingness to support free pedagogical use of their work must be included. The same general rules for evaluation will apply to all electronic submissions. All submissions will be evaluated for (1) pedagogical content, (2) clear description of goals and expected student outcomes, (3) transferability to other settings, (4) appropriateness for the target audience, and (5) references to related educational literature. After publication of the electronic work(s), the authors will be encouraged to submit the work to other databases (National Digital Library, BEN, etc.); however, it is expected that the LSE publication citation will remain associated with the work. This will allow viewers to read a more in-depth discussion of the work.

All such electronic works must be freely available, and will be hosted on the LSE server or on the ASCB server, with the exception of large databases. This will ensure stable access to the works with a nonchanging URL.

A manuscript should accompany any electronic submission. The manuscript should describe 1) the learning goals or purpose of the electronic work, 2) the target audience, 3) development of the electronic work (describe hardware and software used), 4) platform availability (see below), 5) a description of any necessary hardware or software, with links to the appropriate sites for downloading (e.g., plugins, helper applications, etc.), and 6) assessment of the work’s impact on student learning.

Ideally, submitted works should work on any platform (PC, Mac, Unix) and on all browsers. If there are known restrictions, these should be included in the manuscript. LSE can help authors test their works for such limitations if they do not have access to certain platforms or browsers. Assessment may come in the form of learning gains (e.g., pre- and posttests) and/or changes in student attitudes toward the electronic work. Length of time spent using the electronic work may be one indicator of student attitudes. If it is possible to measure student responses to online questions, these responses (number of submissions before correct answer, typical errors, etc.) may further a reader’s understanding of how the electronic work could be incorporated into his or her teaching.

The electronic work may have been hosted previously on any website, but the authors may not have previously published any description of the electronic work other than the associated Web pages. Published journal descriptions of the electronic work will preclude publication in LSE, with the exception of abstracts or presentations at professional meetings.

The manuscript should include an abstract and any figures deemed helpful when the published work appears as a PDF file.

The online publication will include hyperlinks to the work that will appear in a new browser window, if appropriate. This capacity could be helpful to the authors since they could provide directions for readers as needed to illustrate particular aspects of the work. The layout and submission process for the manuscript accompanying an electronic work should follow the same general format as other categories. The electronic work should be submitted to LSE at the same time as the manuscript. If this presents a problem, contact LSE staff for assistance.

Sharing Materials and Data

Publication of a manuscript in LSE implies that the authors agree to make available, to the extent legally permissible, all propagative materials such as mutant organisms, cell lines, recombinant plasmids, vectors, viruses, and monoclonal antibodies that were used to obtain results presented in the article. Prior to obtaining these materials, interested scientists will provide the authors with a written statement that they will be used for noncommercial research purposes only. Instructional materials that were developed in the course of a project described in a LSE article should also be made freely available, preferably in electronic format. Authors are encouraged to share raw data with qualified researchers who wish to reproduce or further analyze the authors’ work for noncommercial purposes. Sharing of data on human subjects should be consistent with the conditions of the Human Subjects protocol approved by the authors’ Institutional Review Board (and any other agreements made with the subjects) for the work reported.

How to Submit Manuscripts

Electronic Submission. Authors must submit manuscripts online at www.cellbiologyeducation.org. Specific instructions on how to submit your manuscript are available at the submission site. Authors should submit the manuscript in either .doc or .rtf as two text files, the first containing the cover letter and the second containing the manuscript file. Figures, tables, and Supplemental Material, including Appendixes, must be submitted as individual files. Do not embed figures in the manuscript. If you are submitting a feature that does not contain an abstract or keywords, write “There is no abstract” in the required abstract field and “one,” “two,” and “three” in the keyword fields.

Questions regarding submission guidelines can be directed to: cbe@ascb.org or 301-347-9304.

General Questions

At any stage in the submission process, authors with questions should contact the LSE Editorial Office at 301-347-9304 (phone); 301-347-9350 (fax); cbe@ascb.org; or The American Society for Cell Biology, 8120 Woodmont Avenue, Suite 750, Bethesda, MD 20814-2762.

You are encouraged to contact editorial board members or the Editor-in-Chief by email to discuss submissions to LSE. The LSE website is an additional resource to authors; see earlier issues of LSE for examples of the different types of manuscripts published.

Last updated on 10/30/13.
A searchable collection of images, videos, and animations of cells and cellular functions, freely available to students, educators and the public.

Explore some of our exciting new features:
* interactive illustrations of *E. coli* and *Paramecium*, in addition to the eukaryotic cell and organelles
* delve into the entire library at once using the Pivot viewer

With your optional free account, you can save images for later use and automatically find new images that match your area of interest.

www.cellimagelibrary.org

We welcome image submissions and feedback and encourage you to visit www.cellimagelibrary.org for more information.
ascb supports the
San Francisco
DORA
Declaration on Research Assessment
ascb.org/SFdeclaration.html
Brought to you by the American Society for Cell Biology and run entirely by scientists for scientists.

IMPORTANT

MBoC presents conceptual advances of broad interest within all areas of cell biology.

INSIGHTFUL

Perspectives, Retrospectives, ASCB Award Essays, and other Features probe topics in science, public policy, and careers.

INTERNATIONAL

The Editorial Board includes Associate Editors in twelve countries.

ACCESSIBLE

Each issue is freely available two months after publication. Authors’ manuscripts are posted within weeks of acceptance.

AUTHOR FRIENDLY

MBoC offers fair, rapid, constructive peer review; free color figures; and a 20% discount on page charges for ASCB members.

www.molbiolcell.org
Apply to be a Biology Scholar!

The mission of the Biology Scholars Program is to empower biologists to be leaders in science education reform and catalyze professional societies to sustain undergraduate education reform. The program goals are achieved through a continuum of three independent, yet intertwined, residencies. Each residency is a yearlong commitment comprised of several key training components including two face-to-face meetings, online “homework” assignments, and specialized peer monitoring.

Assessment Residency: Measuring Student Learning

The Assessment Residency seeks biologists who utilize varied teaching strategies in their classrooms and laboratories and are motivated to design courses and assessments to maximize student learning. Competitive applicants are individuals who are asking questions about the purpose of their courses AND the effectiveness of their course assessments.

Training Institute: June 25 - 28, 2014 | Washington, DC

Research Residency: The Scholarship of Teaching and Learning

The Research Residency is the flagship course of the Biology Scholars Program and has eight years of experience training Scholars in education research. This Residency seeks biologists who have been trained in effective teaching strategies for biology education and are curious about student learning. Competitive applicants are individuals who are asking questions about the effectiveness of their teaching approaches.

Training Institute: July 23 - 26, 2014 | Washington, DC

Transitions Residency: From Science Education Research to Publication

The Transitions Residency seeks biologists who are studying learning in classrooms, ready to share their preliminary results, and ready to evaluate whether their work is ready for publication. Competitive applicants should have initiated their research, collected preliminary data, conducted preliminary analysis of their data, and possibly presented their work to colleagues on campus, virtually, or at regional and national meetings. In addition, individuals who have submitted a biology education research manuscript unsuccessfully are encouraged to apply.

Training Institute: July 23 - 26, 2014 | Washington, DC

Application Deadline for Residencies: February 1, 2014

Travel grants are available to biology faculty members and postdoctoral fellows.

www.biollogyscholars.org
biologyscholars@asmusa.org
The **HHMI EXROP** and **Gilliam Fellowship programs** have as their long-term goal the increase in the diversity of the science professoriate. Our strategy is to provide highly talented students the opportunity to work with top scientists, and to support the students as they navigate through their graduate training.

The **Gilliam Fellowships for Advanced Studies program** provides support to selected EXROP alumni who are pursuing their PhD in science. Since 2004, there have been 62 Gilliam Fellows. Fifteen Fellows have now completed the PhD.
The Exceptional Research Opportunities Program (EXROP) matches outstanding undergraduate students with HHMI scientists in a summer-long mentored research experience. In 2013, there were 76 EXROP students, including 23 “capstone” students who returned to their 2012 labs for a second summer of research. Since 2003, EXROP has sponsored 654 students from 143 colleges and universities who worked with 202 HHMI scientists. More than three-fourths of the students are underrepresented minorities. Of the 370 EXROP alumni who have completed the baccalaureate degree, 362 are in science careers including 186 who have progressed to graduate programs.

The 2013 HHMI EXROP Students

To learn more about the HHMI EXROP Program visit www.hhmi.org/exrop
To read about the 10 Years of EXROP check out the HHMI Bulletin at www.hhmi.org/bulletin/exrop-scientist
To learn more about the HHMI Gilliam Fellowship visit www.hhmi.org/gilliam