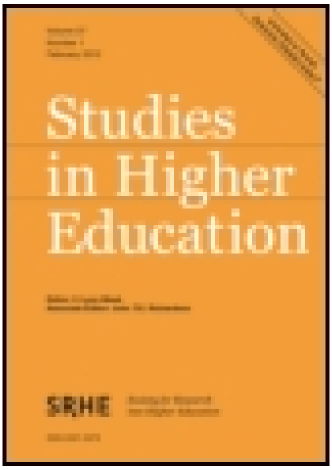

Studies in Higher Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cshe20

Toward a conceptual framework for measuring the effectiveness of course-based undergraduate research experiences in undergraduate biology

Sara E. Brownell[a] & Matthew J. Kloser[b]
[a] School of Life Sciences, Arizona State University, Tempe, AZ, USA
[b] Institute for Educational Initiatives, University of Notre Dame, Notre Dame, IN, USA

Published online: 03 Mar 2015.

Studies in Higher Education, 2015
http://dx.doi.org/10.1080/03075079.2015.1004234

Toward a conceptual framework for measuring the effectiveness of course-based undergraduate research experiences in undergraduate biology

Sara E. Brownell[a*] and Matthew J. Kloser[b†]

[a]*School of Life Sciences, Arizona State University, Tempe, AZ, USA;* [b]*Institute for Educational Initiatives, University of Notre Dame, Notre Dame, IN, USA*

Recent calls for reform have advocated for extensive changes to undergraduate science lab experiences, namely providing more authentic research experiences for students. Course-based Undergraduate Research Experiences (CUREs) have attempted to eschew the limitations of traditional 'cookbook' laboratory exercises and have received increasing visibility in the literature. However, evaluating the outcomes of these experiences remains inconsistent and incomplete partly because of differing goals and conceptual frameworks on the part of those both teaching and assessing the courses. This paper synthesizes existing literature on CUREs and assessment practices to propose a framework for how researchers and practitioners may better align the goals and evaluation practices of CUREs in order to provide a more consistent view of these reformed laboratory courses for the field.

Keywords: CURE; course-based undergraduate research experience; inquiry; cookbook; scientific practices; collaboration; faculty

Recent educational reforms in the USA have emphasized the importance of maintaining and expanding our nation's intellectual resources in the fields of science, technology, engineering, and mathematics (STEM) (e.g. American Cancer Society, Burroughs Wellcome Fund, and Howard Hughes Medical Institute (HHMI) 2000; Obama 2013; President's Council of Advisors on Science and Technology (PCAST) 2010, 2012). To address these needs, several national panels have advocated for changes to undergraduate science education to engage students in more authentic experiences in the content and practices of science (Brewer and Smith 2011; National Research Council (NRC) 2003; PCAST 2012). As citizens, these experiences will help students increase their scientific literacy – the conceptual understanding and knowledge necessary to make informed decisions about the natural and material world (DeBoer 2000; Uno and Bybee 1994). As potential scientists, these experiences will help students improve their understanding of the complex and often misrepresented nature of science (Lederman 1992). Currently, most research experiences are only available to a small, select population of students who get the opportunity to work in faculty

*Corresponding author. Email: sebbers@gmail.com
†Sara E. Brownell and Matthew J. Kloser contributed equally.

research labs; there is a need for colleges to engage a larger, potentially more diverse, student population (Bangera and Brownell 2014).

Re-envisioning the role that laboratory courses and practical work play in undergraduate science education provides one avenue for expanding access to authentic research experiences. The National Academies' National Research Council (2003) report *Bio2010* recognized the need for engaging students in authentic research, stating that 'research with undergraduate students is in itself is the purest form of teaching' (NRC 2003, 87). A more recent national report, *Vision and Change: A Call to Action,* articulates the position of over 500 biologists and biology educators in the USA, arguing that 'cookbook'-type labs – that is, labs in which students follow procedures like a recipe, often without understanding the purpose or methods of the investigation – should be replaced by courses that engage students in learning experiences that better reflect the nature and practices of authentic scientific research (Brewer and Smith 2011). Furthermore, members of a Presidentially-appointed panel in the USA raised the issue of providing more effective lab experiences as one of its five overarching recommendations. The panel argued that undergraduate science students should all 'be given the opportunity to generate scientific knowledge through research' (PCAST 2012, 25), citing that authentic research experiences have been shown to increase STEM retention, especially for traditionally under-represented populations in the sciences (Carter, Mandell, and Maton 2009; Hippel et al. 1998; Russell, Hancock, and McCullough 2007). The message from these reports is clear – undergraduate science students should participate in knowledge-building research, an activity that currently occurs in college for some students, but rarely occurs at scale or at an introductory level.

In response to these calls for reform, some colleges and universities have integrated course-based undergraduate research experiences (CUREs) into their curriculum (Auchincloss et al. 2014). While the courses are diverse in format and organization, many share a common goal of teaching students how to do science through longitudinal investigations of authentic scientific questions as a for-credit class. Despite published articles on CUREs (e.g. Harrison et al. 2011; Weaver, Russell, and Wink 2008), the impact of these CUREs is still in question. This paper builds on existing research and our own experience in assessing authentic research-based lab courses to develop a 'meta-framework', that is, an overarching CURE assessment framework, that can be applied across various CURE iterations, as a means to provide the field with more consistent and comparable information about the impact and best practices of these course-based research experiences. First, we detail the differences between cookbook lab-based biology courses, inquiry-based lab courses, and more authentic CUREs. Next, we build on the work of a National Science Foundation-sponsored panel that has defined five critical components of CUREs and discuss assessments aligned to these five dimensions (Auchincloss et al. 2014). In doing so, we reflect on our own practice of assessing introductory biology CUREs. Finally, we propose an overarching model – the meta-framework – for assessing CUREs by prioritizing aspects of assessment that while challenging and resource intensive, will benefit the field in better understanding how to best prepare undergraduates in rigorous and effective ways for the enterprise of science.

A continuum of undergraduate lab experiences

> Learning science means learning to do science. – Vision and Change: A call to action. (Brewer and Smith 2011)

Science lab courses are often characterized as confirmatory exercises in which students perform tasks that produce a known answer and can be graded as right or wrong (Abrahams and Reiss 2012; Bruck, Bretz, and Towns 2008; Hofstein and Lunetta 2004; Millar 2004). These 'cookbook' experiences focus on learning techniques and following procedures, which is often at odds with rapid advancements in biology. These cookbook labs give students an inaccurate representation of ways in which science is actually done. Science is not a 'rhetoric of conclusions' (Schwab 1962), rather, it is a messy, creative, social, iterative, and human process (Chalmers 1986; Collins and Pinch 1993; Kuhn 1962). Anthropological accounts of science depict lab work not in terms of the execution of techniques, but rather as participants' constant engagement with inscriptions – reading new literature, analyzing and interpreting data, making claims, justifying claims, revising previous work, and constantly discussing, writing, and annotating text (Latour and Woolgar 1986). The linear set of instructions and the emphasis on getting the 'right' answer in cookbook lab courses gives students the impression that science is a collection of facts that we already know and a set of procedures that we need to follow step by step. Furthermore, the cookbook format reduces the natural, collaborative engagement of an authentic lab by assigning data analysis to individual students for homework. Students do not engage with the broader scientific community of their classroom, missing out on opportunities for critique that are paramount to scientific objectivity (Longino 1990).

Recommendations over the past two decades have increasingly emphasized the need to shift away from cookbook toward inquiry-based labs in which students may be given some freedom in defining procedures and analyzing data (e.g. Luckie et al. 2004; Park Rogers and Abell 2008; Sundberg and Moncada 1994). The National Research Council (1996) defines inquiry as 'activities of students in which they develop knowledge and understanding of scientific ideas, as well as an understanding of how scientists study the natural world' (NRC 1996, 23). Evidence from K-12 settings suggests a positive relationship between inquiry science instruction and student outcomes (Furtak et al. 2012).

Shifts toward more inquiry-based labs are also evident among the nation's public and private universities. According to a survey of 118 institutions, only 10% of colleges used what was perceived as an inquiry-based laboratory curriculum in 1993 while over 70% of colleges reported using inquiry-based laboratory instruction in 2005 (Sundberg and Armstrong 1993; Sundberg, Armstrong, and Wischusen 2005). Some studies have suggested, however, that most undergraduate laboratory activities can still be categorized holistically as cookbook in nature (Buck, Bretz, and Towns 2008). This may be due to misinterpretations of what constitutes 'inquiry-based' courses; Brown et al. (2006) interviewed college biology professors and came to the conclusion that 'the overriding constraint to implementing inquiry among faculty was not the logistical, nor even the perceived student factors, but the instructor's meaning of inquiry' (Brown et al. 2006, 798).

Based on the work of Brown et al. (2006) and Buck, Bretz, and Towns (2008), the inquiry classification shown in Table 1 illustrates the varying levels of independence for different scientific practices afforded to students in different lab settings. On the one end of the continuum, confirmation/cookbook labs severely limit the autonomy and decision-making of students, thus drawing the extensive criticisms cited above. On the other end, authentic inquiry is elevated as the highest form of engagement because it works to construct new knowledge. However, this distinction has its own problems; the student autonomy often correlated to open and authentic labs does not necessarily result in effective educational experiences for students. Students may

Table 1. Inquiry classification scheme.

		Presence or absence of student independence by scientific practice						
	Inquiry lab type	Research question	Theoretical background	Methods	Analysis	Conclusions	Communication of results	Known Answer
Cookbook	Confirmation	NO	NO	NO	NO	YES	YES	YES
CUREs	Structured	NO	NO	NO	YES	YES	YES	YES
CUREs	Guided	NO	NO	YES	YES	YES	YES	YES
CUREs	Open	NO	YES	YES	YES	YES	YES	NO
CUREs	Authentic	YES	YES	YES	YES	YES	YES	NO

Source: Adapted from Buck (2008).

initially need developmentally appropriate exercises that target specific scientific practices and aspects of the nature of science. For example, introductory college students may be able to learn much more about the process of science from an experiment where the focus is exclusively on data analysis, whereas more advanced students may benefit more from the experience of asking their own question.

We have adapted this table to include how cookbook labs and CUREs could fit into this scheme: (1) cookbook labs are the same as confirmatory labs in Buck's hierarchy and (2) CUREs have the potential to align with structured, guided, open or authentic inquiry because CUREs focus more on the nature of the scientific practices than a singular level of inquiry, thus calling into question whether a hierarchical inquiry continuum is the best reflection of student learning.

Focusing on research skills and scientific practices, rather than ambiguous notions of inquiry, better aligns students' experiences with the attributes that scientists value – to engage in the excitement, creativity, and unknown exploration of the natural world. Many undergraduates are already doing so in independent research experiences in campus-wide independent research labs. However, there are simply not enough positions at Research 1 institutions and non-research institutions have little to no capacity to reach their students.

In response, CUREs have been created as a way for students to experience authentic research in the context of a lab course that serves students at scale. In some instances, CUREs are extensions of faculty members' research programs – classes are based on a faculty member's research expertise allowing students to collect and analyze data that then might inform the faculty's own research. Other CUREs are stand-alone courses where students work on authentic research problems with unknown answers that do not directly tie to a faculty member's research program. More innovative models also exist; some CUREs engage students from multiple institutions who collaborate on a common problem (e.g. HHMI-funded Science Education Alliance - Phage Hunters Advancing Genomics and Evolutionary Science (SEA PHAGES), Jordan et al. 2014). Regardless of the format, CUREs shift the focus of lab courses from cookbook labs to more authentic research experiences in which students work with a sustained, authentic problem using scientific practices and ways of thinking.

Aligning assessments and the critical components of CUREs

> Although it may seem to be the most obvious way to assess the quality of undergraduate education, the use of direct measures of student learning is uncommon. Marc Chun (2002)

The proliferation of CUREs is seen by many as an exciting pedagogical shift towards experiences that can better prepare the next generation of scientists. As a relatively new phenomenon, the question of how to most effectively assess these courses is still under consideration. The following section draws on a recent publication that identifies five critical components of CUREs (Auchincloss et al. 2014). We discuss each construct in depth and describe assessments that directly or indirectly measure these outcomes. To provide context and dimension to our proposed assessment framework, we draw on our own case study of lessons learned from assessing and evaluating two introductory biology CUREs at a research institution. More information, including specific information about the CUREs, instruments, and assessment items from the courses that we assessed, can be found in Kloser et al. (2011, 2013) and Brownell et al. (2012, 2013).

CURE critical components (Auchincloss et al. 2014) and assessment alignment

1. *The use of science practices*

Defining the use of science practices

The goals of undergraduate science education include not just the acquisition of content knowledge, but also the opportunity for students to successfully engage in the practices and processes of science (Brewer and Smith 2011; NRC 2003; PCAST 2012). Developing future scientists requires opportunities for students to engage the epistemic elements of science – the methods and discourses of justification. In many K-12 settings, a focus on understanding the nature and epistemology of science results in a formulaic application of the five-step scientific method (Rudolph 2005a, 2005b). This standardized version of the 'scientific method' fails to recognize the multiple approaches scientists take to construct new knowledge.

Recent literature has focused on science as a set of scientific practices that are used in concert with core disciplinary ideas and epistemic approaches to explore questions about the natural world (Duschl 2008; NRC 2012). A scientific practice goes beyond a singular technique. Rather, a social practice like that found in the lab is a more or less coordinated, patterned, and meaningful interaction of people at work that makes sense only within reference to a particular activity system (Reckwitz 2002). Applied to a laboratory setting, the understanding of science as practice suggests that common elements exist among members of an system that are central to the nature of the work – for instance, the construction and evaluation of theoretical models, the analysis and interpretation of data, the development of evidence-based arguments and the communication of those arguments represent foundational scientific practices. To be an authentic part of the scientific community, one must engage in these practices and understand the norms for the practices that have been established. Thus, science practice moves individuals beyond the execution of simple techniques; a focus on practice combines knowledge of and about science, routines and skills, and scientific ways of thinking to generate new knowledge about the natural and material world. For the purposes of this paper, we organize scientific practices into three general themes: (1) thinking like a scientist; (2) communicating like a scientist and (3) using the tools of scientists.

Thinking like a scientist focuses on the set of cognitive processes that requires declarative, procedural schematic, and strategic knowledge types (Shavelson, Ruiz-Primo, and Wiley 2005). When thinking like a scientist, students must know disciplinary ideas deeply, know how these ideas are connected and why they are important, and know when, where, and how to use this knowledge to accomplish a task. Indeed, expert scientists are able to use their deep and well-structured content knowledge (Chi, Feltovich, and Glaser 1981) to raise relevant questions, analyze and interpret patterns in data, and construct arguments from evidence. Similarly, the design of experiments and investigations requires the procedural knowledge of how to incorporate positive and negative controls, the schematic knowledge for proper sampling, and the strategic knowledge to evaluate possible flaws in the design.

Communicating like a scientist encompasses the appropriate discourse practices used before, during, and after an investigation. As part of the scientific enterprise, participants are responsible to the local and broader community to report findings orally and in writing. Standard practice assumes that participants in science explicitly and precisely capture their thoughts, procedures and data. Ultimately, this mass of information is condensed and communicated to others in the scientific community following existing norms. For example, the communication of results is necessarily supported by evidence and connected through warranted reasoning (Latour and Woolgar 1986; Toulmin 2003).

Using the tools and technology of science powerfully expands one's ability to capture data. While thinking like a scientist addresses how an individual selects the tools and technologies most appropriate to provide evidence for the claims in question, authentic science engagement also requires attention to precision (e.g. calibrating a mass spectrometer), physical skill (e.g. the proper pipetting of reagents), and the procedural knowledge of standard tests (e.g. running a PCR). In many current cookbook lab courses, this final category comprises the primary aim of student experience as they follow a given set of procedures and develop competencies in basic skills.

Aligning assessment to scientific practices

Incorporating every scientific practice within a single CURE is unfeasible. Rather, 'the opportunity to engage in multiple (e.g. not only data collection) – not all – scientific practices is a CURE hallmark' (Auchincloss et al. 2014, 31). We advocate that while not all scientific practices can be effectively experienced or assessed in one CURE, it is possible to incorporate and assess instances of each of the three categories described above.

Thinking like a scientist. Assessing student progress towards thinking like a scientist can be measured both in terms of competencies and in terms of students' confidence and interest in the tasks. In our evaluation of CUREs, we administer a Likert-scale pre- and post-survey that probes student confidence about their scientific thinking abilities. While we found gains in student confidence in their ability to design an experiment and analyze data (Brownell et al. 2012; Kloser et al. 2013), we do not know from this data whether students are better at these skills or if they merely *think* that they are better. This is a similar problem with data from the commonly used *Classroom Undergraduate Research Experiences Survey* (Lopatto and Tobias 2010); it only measures students' self-reports of their ability to conduct research, a measure of perceived competencies, not competencies themselves. As some studies have shown a relationship between self-confidence and achievement (e.g. Bandura 1997; Bryan, Glynn, and Kittleson 2011), we do not suggest that these affective variables be ignored. Rather, we strongly advocate the addition of direct measures of students' competencies.

While there are a number of previously validated instruments to generally measure experimental design abilities, including the Experimental Design Ability Tool (Sirum and Humburg 2011), the Expanded-Experimental Design Ability Tool (Brownell et al. 2014), the Rubric for Experimental Design (Dasgupta, Anderson, and Pelaez 2014), and a biological experimental design concept inventory (Deane et al. 2014), they provide scenarios and contexts that are far removed from the context of the lab course. As expert thinking is tied to deep content knowledge, we advocate that measurements of scientific thinking should present students with transfer tasks within the existing scientific domain (e.g. the same model system explored during the CURE). An appropriate task might introduce a new variable to the system explored, ask students to design an investigation to address a new research question, and provide previously unseen data from which the students must make and justify initial claims. As part of our own work, we designed a pre- and post-course performance assessment with isomorphic prompts that asked students to design an experiment and analyze data (Kloser et al. 2013). This near transfer task asked students to consider data similar to data that they were working within the course, making it a novel task but closely aligned with the work they had done in the course. Developing transfer tasks should use the steps of the 'assessment square' (Ayala et al. 2002) to logically decompose the construct, develop proper items, pilot the items using think alouds, and revise the items in response to the cognitive analysis. This approach, while time and resource intensive, is preferable in our opinion to using existing instruments that decontextualize the assessment items.

Communicating results. Students' ability to communicate results can be assessed with authentic products such as written manuscripts structured as an academic journal article or through a conference-like poster presentation. In both instances, students must capture the research holistically and interact with colleagues through the answering and posing of questions. Students are not expected to produce a journal quality paper in the brevity of a single term; however, the clarity of communication, the coherence, and the ability of the student to articulate an evidence-based claim can be measured with a rubric against established criteria and performance indicators.

In one of the CUREs that we assessed, students presented a poster in a conference-like environment. Student posters were evaluated using a rubric that identified elements of a poster that are standard in scientific presentations. While these students felt as though their communication skills had improved on Likert-scale post-course surveys compared with pre-course surveys, a more direct analysis of their communication skills is needed before we can conclude that the course benefited their communication skills.

Using tools and technology. Given the broad range of scientific practices that could be assessed, we contend that assessment priorities for CUREs should lie with the first two categories – thinking and communicating like a scientist. While lab practicals may be used to measure students' physical manipulation of tools, the variability and rapidly changing nature of technology make the long-term consequences of assessing the purely procedural, skill-based practices less meaningful. Monitoring of students' use of tools should occur formatively, allowing faculty instructors and teaching assistants to make instructional decisions about which skills to model, coach, or re-teach because proper use is needed to achieve accurate results. However, this formative assessment should not divert needed resources away from measuring students' ability to think and communicate like a scientist.

Summary of aligning assessments for scientific practices. Overall, scientific practices – ways of thinking, communicating, and using tools and technology – require multiple forms of assessment, human capital, and input from both the course instructional team and external evaluators. Just focusing on traditional measures of students' cognitive outcomes or their ability to accomplish procedural tasks falls short in measuring the holistic and relevant nature of CUREs. The remaining four critical components of CUREs raise questions about constructs that, while difficult to assess, are essential to fully understanding the impact of CUREs.

2. *Collaboration*

Defining collaboration

Collaboration and social interaction have great potential to enhance student learning and affect (Johnson and Johnson 1986; Johnson, Johnson, and Smith 1991, 1998). A collaborative environment has been shown to positively affect student learning when compared to passive, active, and even constructive learning environments (Chi 2009). Unfortunately, even in courses designed to foster interactivity, the quality of peer-to-peer social engagement for the purposes of learning is often lacking. For example, one study in a lecture-based science class that dedicated time to students discussing conceptual problems found that only 36% of students' statements were interactive in nature (Willoughby and Gustafson 2009).

While truly interactive environments may be rare in undergraduate science courses, CUREs must incorporate collaborative opportunities if they are to truly reflect authentic research experiences. Collaboration in lab courses must move beyond students working side-by-side to students working interactively to accomplish a particular task; one would not see a university research facility in which two-dozen student pairs worked side-by-side on the same task without contributing data, insight, and expertise to a shared problem. Authentic research is increasingly interdisciplinary in which contributors bring specialized skills and knowledge to the lab (NRC 2003; PCAST 2012), divide tasks and share results, and through collaborative discourse and critique, move the work forward.

Many cookbook and inquiry-based labs use a model where students collect the data together in partners and then analyze and interpret the data individually, outside of class time, for the creation of a lab report. However, in some CUREs, students not only collect the data together, but they also analyze, interpret, and present the data to their colleagues. Additionally, it is unlikely that lab partners can collect enough data by themselves to make claims about unknown phenomena; collaboration and interactivity can be fostered by creating organizational structures in which students divide data collection amongst themselves, multiple groups do the same experiment for increased replicates, and students share data across lab partners, lab sections, and even among different years of the course to present longitudinal findings. This form of collaboration around data better reflects authentic research approaches and requires students to be accountable to colleagues.

While collaboration is a critical component defined by the Auchincloss paper, we argue that interactivity, which includes the type of collaboration mentioned above, is the actual construct of interest. CUREs must provide opportunities for extensive science discourse and critique (Alexander 2005; Lemke 1990). Therefore, these lab courses should include dedicated opportunities for students, familiar with the larger data set, to raise questions about the ideas and hypotheses raised by other students, how they analyze data, and how they justify their claims.

Aligning assessment to collaboration

Assessing collaboration at the student level can be difficult. Existing instruments for discussion and critique are available, including, the Communicative Interactions section of the Reformed Teaching Observation Protocol (Sawada et al. 2002), the Inquiring into Science Instruction Observation Protocol (Minner and DeLisi 2012), and the Discourse in Science Inquiry Classrooms tool (Baker 2008). These instruments, while not designed for CUREs, can be used as rubrics to score the type of discourse and interactions that occur in class or during summative presentations by looking at students' level of questioning and the uptake, elaboration, and clarification of student ideas during discussion.

Collaboration can also be evaluated through curricular analysis of lesson plans and the syllabus by asking: Do students have opportunities to collect different data than other students in their class? Do students contribute data and findings to a centralized database for use by peers? Do students have opportunities to engage other lab pairs in discussion about possible findings? Second, informal student interactions can be sampled and analyzed for the level of discourse that occurs in peer-to-peer interactions. Pairs should be strategically sampled and recorded or observed to determine the types of activities that occur.

We assessed the level of collaboration and interactivity in a CURE by collecting observational data from a cookbook course and a CURE. An observation instrument was created and calibrated by the authors of this paper and data was recorded for the presence of particular individual, collaborative, and discursive behaviors (e.g. listening to a lecture, asking and answering questions, off-task behavior) each minute in an eight-minute interval. Currently unpublished results reveal that students in the CURE were more likely to engage in interactive and more sophisticated discursive practice, at higher cognitive Bloom's levels, including application, analysis, evaluation, and synthesis. We videotaped students' summative presentations and took field notes during the live sessions. Qualitative observations indicated significant amounts of critique in which students respectfully challenged each other's analysis and raised questions about future directions of research; formal use of a rubric that captures levels of uptake, elaboration, and critique would allow us to make more specific claims about how the course's structure influences the critical component of collaboration.

Regardless of the instrument chosen, assessing collaboration and interactivity should also attend to issues of discourse equity. Individuals can dominate in collaborative settings at the expense of other students. Instruments should account for not only the types of discourse, but also who is talking, and how often. For formative purposes, students may individually complete survey items in which they reflect on their own collaborative and discursive participation and those of their partners. This data can provide in-time feedback to instructors about social dynamics that are working or those that are not working and may need to be addressed. This could lead to great insights into how to structure lab groups, including partner pairs, to promote equity.

3. *Iteration*

Defining iteration

Iteration is the repetition of a process, idea or task; it is an integral part of scientific research. Experiments must be repeated to determine the reliability and replicability of the results. There is variability in all biological systems that could lead to different results, depending on the sampling. Thus, iteration includes repeating experiments, refining ideas, and building on those ideas with different types of experiments. In a CURE, 'students may also build on and revise aspects of other students' investigations, whether within a single course to accumulate a sufficiently large data set for analysis or across successive offerings of the course to measure and manage variation, further test preliminary hypotheses, or increase confidence in previous findings' (Auchincloss et al. 2014).

Understanding the need for replication and iteration – and at what amount – is central to a student becoming a part of the scientific enterprise. Exposing students to the nature of science means that they understand the important role that iteration, not one-time experiments or observations, plays in justifying one's own claims (McComas 1998; Osborne et al. 2003).

Aligning assessment to iteration

Assessing student iteration of experiments is important to students' understanding of science as an enterprise. Previous studies show that students can hold inaccurate conceptions related to repeating an experiment, which may be due to the lack of iteration in most science courses (Brownell et al. 2014). The inherent variability of data is often hidden if students collect only one set of data or are only given the average observation. Attending to the critical component of iteration can help students learn how to manage failed experiments, both practically in terms of what to do next in the lab and emotionally as far as dealing with overcoming failure. Failed experiments give students opportunities to troubleshoot what might have gone wrong and repeating particular experiments gives them greater insights into sources of error. Assessing students' responses to failed experiments could be difficult, but may be included with other affective components, such as interest and self-efficacy, in a survey. Qualitative measures may also be employed wherein students are asked to reflect on their experience at three or four times throughout the course. These embedded prompts could probe students' disposition toward and understanding of the need to replicate experiments and overcome failure. These responses can be coded and charted according to a timescale to see whether students' feelings move away from negativity toward failure and more toward failure as part and parcel to advancing knowledge. These formative prompts can be taken from existing nature of science instruments (e.g. Abd-El-Khalick et al. 2001) that provide students an opportunity to explain their perspectives and understandings of iteration and failure.

4. *Discovery and 5. Broadly relevant work*

Defining discovery and broadly relevant work

Traditional lab courses use experiments and investigations that are good exemplars of phenomena, ensuring, to the greatest extent possible, that students get the 'right' solution. These cookbook labs not only have known answers, but the same experiments are often conducted for several years so that students have access to the answers from other students. This experience fails to provide students with the excitement of discovery tied so closely to scientific research – the understanding that science is a cumulative process and that many individuals must contribute small pieces of knowledge to come to better understandings of larger systems and phenomena. Creating opportunities for students to experience open-ended discovery is central to engaging the next generation of scientists (Brewer and Smith 2011; PCAST 2012).

Similarly, engaging students in authentic tasks of relevance to their own lives or situations important to the world is also central to illuminating to students the value of well-executed scientific research. In some cases, the relevance will derive from regional needs and issues. Biology students may work on investigating ecological or climate data for a nearby wetlands or chemistry students may investigate the most efficient and safe process for neutralizing chemicals that have contaminated local soil. For both critical components – discovery and broader relevance – students need to work within a system in which students *and* faculty do not know the results at the beginning of the investigation; it is important that students cannot reference existing studies to find published descriptions of what they are supposed to be testing. The messiness and ambiguity of real discovery should not be viewed critically, but rather be viewed as part of the scientific enterprise and reason for more investigation. However, instructors need to be comfortable with a level of flexibility in the classroom because research by its very nature is unpredictable.

Aligning assessment to discovery and broadly relevant work

Publications in peer-reviewed journal articles are the currency of authentic research. Publication indicates novel and important findings, synonyms of discovery-based work that has broad relevance. Given that CUREs can focus on the research programs of faculty, one way to assess whether work conducted in a CURE is broadly relevant is to measure the number and quality of peer-reviewed publications that result from a course. However, just because a publication does not result from the course does not mean that the course does not lead to the discovery of new data or that it is broadly relevant. The time and effort it takes to turn student data into a publication may mean that using publications as a metric would underestimate the discoveries occurring in CUREs. Some student-conducted experiments, especially those that require high levels of technical expertise, may not produce a publishable result, even though they are novel. This does not make it any less important or unique. In fact, some research questions may require the combined effort of several years of student-collected data to acquire enough warrant for a publication.

Analyzing syllabi, curricular plans, and course handouts with a well-defined rubric can also illuminate the level of a course's discovery opportunities and relevance. Alternatively, students can be interviewed or surveyed about their perceptions of the authenticity of the work and its broader relevance. These results can be used both formatively for single-course improvement, as well as summatively for communicating the implications of assessment results to other interested parties in the field.

Three scientific publications have resulted from student work done in the CURE we assessed, indicating that the work is novel and broadly applicable in the larger scientific domain (Belisle, Peay, and Fukami 2012; Peay, Belisle, and Fukami 2012; Vannette, Gauthier, and Fukami 2012). We have used these publications as a metric that the CUREs do allow for discovery and broad applicability, but that even without these publications, the organization and structure of the course would reflect the presence of these constructs.

Summary of aligned assessment in dimensions

Table 2 summarizes the five critical components and their aligned assessments as described above. Examples from our assessment of two CUREs at a Research 1 university are included, as relevant. Further exploration of these assessments can be seen in context in the cited papers in the table.

Table 2. Summary of aligned CURE assessments.

Critical components	Aligned assessment(s)	Examples from our work
Scientific practices: thinking like a scientist	Pre/post near-transfer tasks focused on scientific thinking constructs; pre/post self-confidence and interest surveys	Data analysis and experimental design abilities were measured using a content-specific performance assessment that was designed specifically for the CURE (Kloser et al. 2013)
Scientific practices: communicating like a scientist	Poster presentations; conference-style presentations with critique; journal article-type summative writing submission	Rubric to assess student poster presentations (unpublished)
Scientific practices: using the tools of a scientist	Lab practical, hands-on experiments	N/A – this was not a course goal
Collaboration	In-class measures of student interactivity	Observation tool charting frequency of low and high level Bloom's questions/discursive practices among students (unpublished)
Iteration	Syllabus and course description analysis	N/A
Discovery/broadly relevant work	Syllabus and course description analysis; Publications and conference presentations	Three scientific publications have been published on data collected in the course

Situating faculty in an assessment framework

The five critical components of CUREs outlined in Auchincloss and colleagues (2014) justifiably focus on student outcomes and participation structures within a course-based research environment. Students need to develop competencies with scientific practices, to collaborate with peers through differentiated tasks, experience the benefits and frustrations of iterative investigation, and seek the relevance behind their novel explorations. However, CUREs blend aspects of teaching and research – two primary functions of higher education faculty at research institutions. Assessment and evaluation should also consider the inputs and outputs for faculty as well, not in terms of their developing research competencies, but rather their pedagogical and curricular decisions, their engagement with research-based courses, and even their ability to leverage CUREs for further data collection and/or publication with students. We observed a tenure-track faculty member who successfully integrated the development and enactment of a CURE in ways that expanded his own research program and led to co-authored publications (Kloser et al. 2011). Faculty members who use their own research to develop a CURE allows faculty to become more engaged in teaching even if they predominantly hold a research identity (Brownell and Tanner 2012).

Building in evaluations of faculty outcomes with CUREs has been an overlooked construct in much of the literature, but it also must be approached with some caution. Ultimately, CUREs are established for students and outcome measures like faculty publications should be viewed as an added, but not necessary value. One could imagine faculty unintentionally limiting students' opportunities to engage in multiple scientific practices in order to obtain a large quantity of particular data for their own use. For example, students could be asked to run myriad Western blots, producing many replications of the same assay, but be limited in their understanding of how the methods fit within the larger scope of research. It is possible that faculty and student goals could come into conflict with each other, so being aware of these potential conflicts is important. Keeping in mind these possible obstacles to student learning, evaluating CUREs should extend beyond student outcomes and include the impact on faculty as well, not only through course evaluations, but also through surveys of faculty perspectives on teaching authentic research-based courses.

A CURE assessment meta-framework

Scientific research can be decomposed into practices, habits of mind, collaborative activity systems, and goal-directed work – but meeting the demands of recent calls for reform implies a holistic experience that integrates these parts. Above, we articulated and prioritized assessments that align with recently identified goals for CURE development. Below, we propose a holistic meta-framework that integrates the above-discussed pieces in order to capture the research experience for both students and faculty. This framework recognizes that labs will vary greatly among (sub)disciplines and institutions, but at minimum, measuring the same global constructs may better allow for systematic comparisons and warranted claims about CUREs that the field currently lacks. We recognize the value of and necessity for formative assessment to constantly improve the CURE experience; however, the framework focuses on the summative outcomes of the course.

Our framework recognizes that CUREs serve a variety of purposes, requiring a focus on three primary types of outcomes that must be measured to get a holistic perspective on the CURE's impact: (1) course outcomes, (2) student outcomes and (3) faculty outcomes. Course outcomes measure whether the designed course integrates the components essential to defining a CURE through its format, its instruction, its curriculum, and its assessment. Assessing the course outcomes answers the question, 'Does this course provide an experience for students to think, communicate, and practice like a scientist?' Student outcomes are clearly tied to the course outcomes in many ways, but measurement of these outcomes focus on the student experience as a learner and contributor. Assessing students outcomes answers the questions, 'What science do students learn? What do they learn about the nature of science? What is their level of interest and self-efficacy in scientific research?' Finally, the measurement of faculty outcomes focuses on measuring benefits or detriments to faculty members' careers. Assessing faculty outcomes helps answer the questions, 'How did this experience affect faculty perceptions of teaching? How did this course advance or impede faculty research programs? What benefits exist from teaching this course that go beyond student learning?'

Figure 1 depicts a Venn diagram of the three domains – course outcomes, student outcomes, and faculty outcomes – that we argue are essential to holistically gathering data and making claims about the impact of CUREs. One cannot collect data on only one of these outcomes and then conclude something about the CURE as a whole. As shown by the numbered and italicized phrases, an effective CURE must have some level of 'face validity' in that the course is structured to address the five major critical components that define CUREs; students must have opportunities to think, communicate, and practice like scientists. Currently, most published evaluations focus solely on student outcomes, often survey-based outcomes that do not measure students'

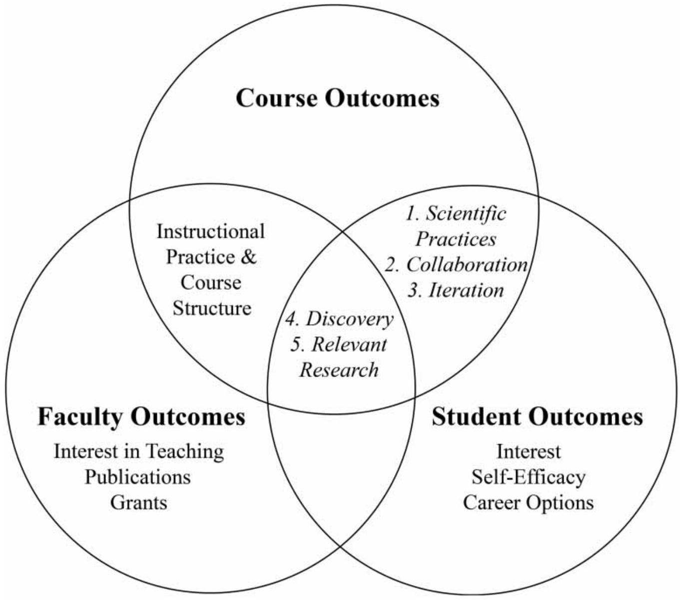

Figure 1. A CURE assessment meta-framework. Critical components are shown in italics while other, related assessment domains are shown in plain font.

engagement in scientific ways of thinking, communicating like a scientist, or collaboration. Affective outcomes, particularly those that have been shown to mediate future engagement, such as student interest, should be measured, but only in concert with measurements of their competencies in scientific practices. Furthermore, faculty teaching practice and benefits to their research are not simply bi-products, but can be fundamental parts of the CURE.

As the framework shows, a majority of the outcomes that need to be measured address multiple domains, satisfying different outcomes. Competencies such as students' ability to interpret data, develop a hypothesis, or communicate results can be assessed from two perspectives – as a course outcome, whether the intended curriculum fosters student engagement with these scientific practices, and as a student outcome, the levels at which students actually performed these tasks and exhibited scientific thinking. Iteration also overlaps between course and student outcomes. The course documents should provide evidence of opportunities for iteration while assessment measures, like the above-mentioned nature of science instruments, should assess how iterative cycles of experimentation and analysis affect students' perceptions of overcoming failure while doing research.

Similarly, measurements of the CURE's intended and enacted pedagogy reflect course and faculty outcomes, respectively. Through self-reflection, course evaluations, and live observations, course and faculty outcomes focused on curriculum and instruction can be measured. This data will be put in even broader context when paired with the outcomes unique to the faculty members – how this lab teaching experience influences their own identity as a teacher and researcher.

Finally, the central section of the Venn diagram represents constructs at the intersection of course, student, and faculty outcomes. Regardless of the CURE, assessments should capture whether the course, students, and faculty engage in lab work that is unique, lab work that seeks to construct new knowledge, and lab work that is relevant within a larger problem space. Developing a course that has a central focus of creation of new and relevant knowledge influences all of the other outcomes and can lead to products, such as conference presentations, journal articles, or new questions for exploration, not possible in a cookbook lab format.

As CUREs proliferate, a complete understanding of their impact requires a suite of measurements that go well beyond single measurements of a student's interest or self-efficacy level. Summative assessments should, at minimum, provide data for whether the CURE meets each of the five critical components (numbered and in italics in Figure 1), but we argue that this does not fully represent the potential value of the course. Faculty outcomes such as publications, grants, and increased interest in teaching and student outcomes such as interest and self-efficacy should be measured along with student competencies. Results from these assessments can then be used formatively for improving individual courses and by following this meta-framework, can contribute to the field's better understanding of whether CUREs better help students meet the goals of intended reforms not met by traditional cookbook lab experiences.

CURE assessment: it takes a village

The proposed assessment framework requires multiple measures and extensive time in data collection, analysis, and the reporting of findings. For pedagogical and measurement reasons, we argue that the most effective assessment plan involves a collaboration between faculty instructors, teaching assistants, departmental administrators, and external evaluators. Faculty members will naturally collect summative data on several student competencies. They can work with an external evaluation team to choose an effective rubric, for instance, of how well students communicate or execute other scientific practices. An external evaluation team provides some level of detachment from the instructional planning and provides human capital to evaluate aspects of the program – like collaboration and interaction – that are difficult to systematically capture when one is also facilitating the course. The external evaluation team also provides opportunities for assessments that are blind to faculty and students, such as pre/post transfer tasks, reducing bias in the process. Departmental leadership is also essential for resource allocation and logistical considerations (e.g. permitting randomization among lab sections when comparing traditional labs and CUREs).

Working collaboratively may blur some aspects of objectivity. In our own experience as external evaluators we developed strong working relationships with the instructional team; by working collaboratively we were able to collect richer and larger amounts of data, balancing the more short-term instructional needs with longer-term evaluation of the program.

Closing thoughts and future directions

Movements across higher education to reform traditional science lab courses are both exciting and necessary. Many aspects of our culture, environment, economy, health, and standard of living are tied to the products of scientific knowledge building. Like other countries, the USA benefits from creating a consistent pipeline of young, innovative minds who actively seek to participate in the practices and language of science. As discussed above, traditional organization and structure of undergraduate science labs do not generally foster these dispositions. CUREs offer a refreshing alternative to cookbook labs since they are more authentic research experiences with potential benefits for both students and faculty members.

Of course, the benefits cannot be assumed. Similar to previous movements to infuse active learning in science classrooms, execution of such changes can occur effectively or ineffectively. Merely changing curriculum and instruction is not enough to warrant widespread changes. Significant resources should be committed not only to the development of feasible and accessible research model systems, but also to assessing student outcomes and competencies, faculty pedagogy, and faculty benefits. In order to better understand the impact of CUREs, the field needs to adopt more common assessment approaches aligned to the five critical components defined by Auchincloss et al. (2014), allowing the aggregation of findings across institutions and time. Furthermore, as traditional lab courses are transformed into CUREs, we encourage a two to three-year phase-in period that allows for randomization of students into lab conditions and the ability to make causal claims about student outcomes. Randomization will allow for important comparisons and using pre/post assessments will also allow evaluators to measure overall growth in both conditions. Only then will the field be able to more confidently understand the impact of these holistic reforms.

Acknowledgements

We would like to thank the CURE instructional teams, particularly Tad Fukami, Martha Cyert, Tim Stearns, Daria Hekmat-Scafe, Shyamala Malladi, and Pat Seawell, as well as the students who participated in the courses. We would also like to thank Bob Simoni and Rich Shavelson.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Funding for the lab design was provided by a [NSF TUES award 0941984].

References

Abd-El-Khalick, F., N. G. Lederman, R. L. Bell, and R. S. Schwartz. 2001. "Views of Nature of Science Questionnaire (VNOS): Toward Valid and Meaningful Assessment of Learners' Conceptions of Nature of Science." In *Proceedings of the 2001 Annual International Conference of the Association for the Education of Teachers in Science*, edited by P. A. Rubba, J. A. Rye, W. DiBiase, and B. Crawford, 212–77. Pensacola, FL: Association for the Education of Teachers in Science.

Abrahams, I., and M. J. Reiss. 2012. "Practical Work: Its Effectiveness in Primary and Secondary Schools in England." *Journal of Research in Science Teaching* 49 (8): 1035–55.

Alexander, R. 2005. *Towards Dialogic Teaching*. York: Dialogos.

American Cancer Society, Burroughs Wellcome Fund, & Howard Hughes Medical Institute. 2000. "The Role of the Private Sector in Training the Next Generation of Biomedical Scientists." http://www.bwfund.org/sites/default/files/media/files/Training%20the%20Next%20Generation.pdf

Auchincloss, L. C., S. L. Laursen, J. L. Branchaw, K. Eagan, M. Graham, D. Hanauer, G. Lawrie, et al. 2014. "Assessment of Course-Based Undergraduate Research Experiences: A Meeting Report." *CBE-Life Sciences Education* 13 (1): 29–40.

Ayala, C. C., Y. Yin, R. J. Shavelson, and J. Vanides. 2002. "Investigating the Cognitive Validity of Science Performance Assessment with Think Alouds: Technical Aspects." Presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Baker, D. 2008. “Discourse in Inquiry Science Classrooms (DiISC): Reference Manual.” http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1120&context=teachlearnfacpub
Bandura, A. 1997. *Self-Efficacy: The Exercise Of Control*. New York: Freeman.
Bangera, G., and S. E. Brownell. 2014. “Course-Based Undergraduate Research Experiences Make Research More Inclusive.” *CBE Life Sciences Education* 13 (4): 602–6.
Belisle, M., K. G. Peay, and T. Fukami. 2012. “Flowers as Islands: Spatial Distribution of Nectar-Inhabiting Microfungi among Plants of *Mimulus aurantiacus*, a Hummingbird-Pollinated Shrub.” *Microbial Ecology* 63: 711–18.
Brewer, C. A., and D. Smith. 2011. *Vision and Change in Undergraduate Biology Education: A Call to Action*. Washington, DC: American Association for the Advancement of Science.
Brown, P. L., S. K. Abell, A. Demir, and F. J. Schmidt. 2006. “College Science Teachers’ Views of Classroom Inquiry.” *Science Education* 90 (5): 784–802.
Brownell, S. E., M. J. Kloser, T. Fukami, and R. Shavelson. 2012. “Undergraduate Biology Lab Courses: Comparing the Impact Of Traditionally Based ‘Cookbook’ and Authentic Research-Based Courses on Student Lab Experiences.” *Journal of College Science Teaching* 41 (4): 36–45.
Brownell, S. E., M. J. Kloser, T. Fukami, and R. J. Shavelson. 2013. “Context Matters: Volunteer Bias, Small Sample Size, and the Value of Comparison Groups in the Assessment of Research-Based Undergraduate Introductory Biology Lab Courses.” *Journal of Microbiology & Biology Education* 14 (2): 176–82.
Brownell, S. E., and K. D. Tanner. 2012. “Barriers to Faculty Pedagogical Change: Lack of Training, Time, Incentives, and … Tensions with Professional Identity?” *CBE-Life Sciences Education* 11 (4): 339–46.
Brownell, S. E., M. P. Wenderoth, R. Theobald, N. Okoroafor, M. Koval, S. Freeman, et al. “How Students Think About Experimental Design: Novel Conceptions Revealed by In-Class Activities.” *BioScience* 64 (2): 125–37.
Bruck, L. B., S. L. Bretz, and M. H. Towns. 2008. “Characterizing the Level of Inquiry in the Undergraduate Laboratory.” *Journal of College Science Teaching* 38 (1): 52–58.
Bryan, R. R., S. M. Glynn, and J. M. Kittleson. 2011. “Motivation, Achievement, and Advanced Placement Intent of High School Students Learning Science.” *Science Education* 95 (6): 1049–65.
Buck, L. B., S. L. Bretz, and M. H. Towns. 2008. “Characterizing the Level of Inquiry in the Undergraduate Laboratory.” *Journal of College Science Teaching* 38 (1): 52–58.
Carter, F. D., M. Mandell, and K. I. Maton. 2009. “The Influence of On-Campus, Academic Year Undergraduate Research on STEM Ph.D. Outcomes: Evidence from the Meyerhoff Scholarship Program.” *Educational Evaluation and Policy Analysis* 31 (4): 441–62.
Chalmers, A. F. 1986. *What is this Thing Called Science*? Milton Keynes: Open University Press.
Chi, M. T. 2009. “Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities.” *Topics in Cognitive Science* 1 (1): 73–105.
Chi, M. T., P. J. Feltovich, and R. Glaser. 1981. “Categorization and Representation of Physics Problems by Experts and Novices.” *Cognitive Science* 5 (2): 121–52.
Chun, M. 2002. “Looking Where the Light is Better: A Review of the Literature on Assessing Higher Education Quality.” *Peer Review* 4 (2/3): 16–25.
Collins, H. M., and T. Pinch. 1993. *The Golem: What Everyone should know about Science*. Cambridge, MA: Cambridge University Press.
Dasgupta, A. P., T. R. Anderson, and N. Pelaez. 2014. “Development and Validation of a Rubric for Diagnosing Students’ Experimental Design Knowledge and Difficulties.” *CBE Life Sciences Education* 13 (2): 265–84.
Deane, T., K. Nomme, E. Jeffery, C. Pollock, and G. Birol. 2014. “Development of the Biological Experimental Design Concept Inventory (BEDCI).” *CBE-Life Sciences Education* 13 (3): 540–51.
DeBoer, G. 2000. “Scientific Literacy: Another Look at its Historical and Contemporary Meanings and its Relationship to Science Education Reform.” *Journal of Research in Science Teaching* 37 (6): 582–601.
Duschl, R. A. 2008. “Science Education in Three-Part Harmony: Balancing Conceptual, Epistemic, and Social Learning Goals.” *Review of Research in Education* 32 (1): 268–91.

Furtak, E. M., T. Seidel, H. Iverson, and D. C. Briggs. 2012. “Experimental and Quasi-Experimental Studies of Inquiry-Based Science Teaching: A Meta-Analysis.” *Review of Educational Research* 82 (3): 300–329.

Harrison, M., D. Dunbar, L. Ratmansky, K. Boyd, and D. Lopatto. 2011. “Classroom-based Science Research at the Introductory Level: Changes in Career Choices and Attitude.” *CBE-Life Sciences Education* 10 (3): 279–86.

von Hippel, W., J. S. Lerner, S. R. Gregerman, B. A. Nagda, and J. Jonides. 1998. “Undergraduate Student-Faculty Research Partnerships Affect Student Retention.” *The Review of Higher Education* 22 (1): 55–72.

Hofstein, A., and V. N. Lunetta. 2004. “The Laboratory in Science Education: Foundations for the Twenty-First Century.” *Science Education* 88 (1): 28–54.

Johnson, R. T., and D. W. Johnson. 1986. “Cooperative Learning in the Science Classroom.” *Science and Children* 24: 31–32.

Johnson, D. W., R. T. Johnson, and K. A. Smith. 1991. Active Learning: Cooperation in the College Classroom. Edina, MN: Interaction Book Company.

Johnson, D. W., R. T. Johnson, and K. A. Smith. 1998. “Cooperative Learning Returns to College what Evidence is there that it Works?” *Change: The Magazine of Higher Learning* 30 (4): 26–35.

Jordan, T. C., S. H. Burnett, S. Carson, S. M. Caruso, K. Clase, R. J. DeJong, J. J. Dennehy, et al. 2014. “A Broadly Implementable Research Course in Phage Discovery and Genomics for First-Year Undergraduate Students.” *mBio* 5 (1): e01051–13.

Kloser, M. J., S. E. Brownell, N. R. Chiariello, and T. Fukami. 2011. “Integrating Teaching and Research in Undergraduate Biology Laboratory Education.” *PLoS Biology* 9 (11): e1001174.

Kloser, M. J., S. E. Brownell, R. J. Shavelson, and T. Fukami. 2013. “Effects of a Research-Based Ecology Lab Course: A Study of Nonvolunteer Achievement, Self-Confidence, and Perception of lab Course Purpose.” *Journal of College Science Teaching* 42 (3): 90–99.

Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Latour, B., and S. Woolgar. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Lederman, N. G. 1992. “Students’ and Teachers’ Conceptions of the Nature of Science: A Review of the Research.” *Journal of Research in Science Teaching* 29 (4): 331–59.

Lemke, J. L. 1990. *Talking Science: Language, Learning, and Values*. Norwood, NJ: Ablex Publishing Corporation.

Longino, H. E. 1990. *Science as Social Knowledge*. Princeton, NJ: Princeton University Press.

Lopatto, D., and S. Tobias. 2010. *Science in Solution: The Impact of Undergraduate Research on Student Learning*. Washington, DC: Council on Undergraduate Research.

Luckie, D. B., J. J. Maleszewski, S. D. Loznak, and M. Krha. 2004. “Infusion of Collaborative Inquiry throughout a Biology Curriculum Increases Student Learning: A Four-Year Study of ‘Teams and Streams’.” *Advances in Physiology Education* 28 (4): 199–209.

McComas, W. F., ed. 1998. *The Nature of Science in Science Education: Rationales and Strategies*. Boston: Kluwer Academic.

Millar, R. 2004. “The Role of Practical Work in the Teaching and Learning of Science.” Paper Presented at the “High School Science Laboratories: Role and Vision” Meeting, Board on Science Education, National Academy of Sciences, Washington, DC.

Minner, D., and J. DeLisi. 2012. Inquiring into Science Instruction Observation Protocol (ISIOP). http://isiop.edc.org/

National Research Council. 1996. *National Science Education Standards*. Washington, DC: National Academy Press.

National Research Council. 2003. *BIO 2010: Transforming Undergraduate Education for Future Research Biologists*. Washington, DC: National Academies Press.

National Research Council. 2012. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, DC: The National Academies Press.

Obama, B. H. 2013. “Remarks by the President in the State of the Union Address,” February 12. http://www.whitehouse.gov/thepress-office/2013/02/12/remarks-president-state-union-address.

Osborne, J., S. Collins, M. Ratcliffe, R. Millar, and R. Duschl. 2003. "What 'ideas-about-science' should be Taught in School Science? A Delphi Study of the Expert Community." *Journal of Research in Science Teaching* 40 (7): 692–720.

Park Rogers, M. A., and S. K. Abell. 2008. "The Design, Enactment, and Experience of Inquiry-Based Instruction in Undergraduate Science Education: A Case Study." *Science Education* 92 (4): 591–607.

Peay, K. G., M. Belisle, and T. Fukami. 2012. "Phylogenetic Relatedness Predicts Priority Effects in Nectar Yeast Communities." *Proceedings of the Royal Society B: Biological Sciences* 279: 749–58.

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: Executive Office of the President.

President's Council of Advisors on Science and Technology. 2012. *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*. Washington, DC: Executive Office of the President.

Reckwitz, A. 2002. "Toward a Theory of Social Practice: A Development in Culturalist Theorizing." *European Journal of Social Theory* 5 (2): 243–63.

Rudolph, J. L. 2005a. "Epistemology for the Masses: The Origins of 'The Scientific Method' in American Schools." *History of Education Quarterly* 45 (3): 341–76.

Rudolph, J. L. 2005b. "Inquiry, Instrumentalism, and the Public Understanding of Science." *Science Education* 89 (5): 803–21.

Russell, S. H., M. P. Hancock, and J. McCullough. 2007. "Benefits of Undergraduate Research Experiences." *Science* 316 (5824): 548–49.

Sawada, D., M. D. Piburn, E. Judson, J. Turley, K. Falconer, R. Benford, and I. Bloom. 2002. "Measuring Reform Practices in Science and Mathematics Classrooms: The Reformed Teaching Observation Protocol." *School Science and Mathematics* 102 (6): 245–53.

Schwab, J. J. 1962. "The Teaching of Science as Inquiry." In *The Teaching of Science*, edited by J. J. Schwab and P. Brandwein, 1–104. Cambridge, MA: Harvard University Press.

Shavelson, R. J., M. A. Ruiz-Primo, and E.W. Wiley. 2005. "Windows into the Mind." *Higher Education* 49 (4): 413–30.

Sirum, K., and J. Humburg. 2011. "The Experimental Design Ability Test (EDAT)." *Bioscene: Journal of College Biology Teaching* 37 (1): 8–16.

Sundberg, M. D., and J. E. Armstrong. 1993. "The Status of Laboratory Instruction for Introductory Biology in US Universities." *The American Biology Teacher* 55: 144–46.

Sundberg, M. D., J. E. Armstrong, and E. W. Wischusen. 2005. "A Reappraisal of the Status of Introductory Biology Laboratory Education in US Colleges & Universities." *The American Biology Teacher* 67 (90): 525–29.

Sundberg, M. D., and G. J. Moncada. 1994. "Creating Effective Investigative Laboratories for Undergraduates." *BioScience* 44: 698–704.

Toulmin, S. E. 2003. *The Uses of Argument*. Cambridge, MA: Cambridge University Press.

Uno, G. E., and R. W. Bybee. 1994. "Understanding the Dimensions of Biological Literacy." *BioScience* 44 (8): 553–57.

Vannette, R. L., M. P. L. Gauthier, and T. Fukami. 2012. "Nectar Bacteria, but not Yeast, Weaken a Plant-Pollinator Mutualism." *Proceedings of the Royal Society B: Biological Sciences* 280: 20122601.

Weaver, G. C., C. B. Russell, and D. J. Wink. 2008. "Inquiry-based and Research-Based Laboratory Pedagogies in Undergraduate Science." *Nature Chemical Biology* 4 (10): 577–80.

Willoughby, S. D., and E. Gustafson. 2009. "Technology Talks: Clickers and Grading Incentive in the Large Lecture Hall." *American Journal of Physics* 77 (2): 180–83.