

Feature Meeting Report

Assessment of Course-Based Undergraduate Research Experiences: A Meeting Report

Lisa Corwin Auchincloss,^{*} Sandra L. Laursen,[†] Janet L. Branchaw,[‡] Kevin Eagan,[§] Mark Graham,^{||} David I. Hanauer,[¶] Gwendolyn Lawrie,[#] Colleen M. McLinn,[@] Nancy Pelaez,^{**} Susan Rowland,[#] Marcy Towns,^{**} Nancy M. Trautmann,[@] Pratibha Varma-Nelson,^{††} Timothy J. Weston,[†] and Erin L. Dolan^{*}

^{*}University of Georgia, Athens, GA 30602; [†]University of Colorado–Boulder, Boulder, CO 80309; [‡]University of Wisconsin–Madison, Madison, WI 53706; [§]University of California, Los Angeles, Los Angeles, CA 90995; ^{||}Yale University, New Haven, CT 06520; [¶]Indiana University of Pennsylvania, Indiana, PA 15705; [#]University of Queensland, Brisbane 4072, Australia; [@]Cornell University, Ithaca, NY 14850; ^{**}Purdue University, West Lafayette, IN 47907; ^{††}Indiana University–Purdue University, Indianapolis, IN 46202

The Course-Based Undergraduate Research Experiences Network (CUREnet) was initiated in 2012 with funding from the National Science Foundation program for Research Coordination Networks in Undergraduate Biology Education. CUREnet aims to address topics, problems, and opportunities inherent to integrating research experiences into undergraduate courses. During CUREnet meetings and discussions, it became apparent that there is need for a clear definition of what constitutes a CURE and systematic exploration of what makes CUREs meaningful in terms of student learning. Thus, we assembled a small working group of people with expertise in CURE instruction and assessment to: 1) draft an operational definition of a CURE, with the aim of defining what makes a laboratory course or project a “research experience”; 2) summarize research on CUREs, as well as findings from studies of undergraduate research internships that would be useful for thinking about how students are influenced by participating in CUREs; and 3) identify areas of greatest need with respect to CURE assessment, and directions for future research on and evaluation of CUREs. This report summarizes the outcomes and recommendations of this meeting.

Students can work with the same data at the same time and with the same tools as research scientists. iPlant Education, Outreach & Training Group (2008, personal communication)

INTRODUCTION

Numerous calls for reform in undergraduate biology education have emphasized the value of undergraduate research

DOI: 10.1187/cbe.14-01-0004

Address correspondence to: Erin L. Dolan (eldolan@uga.edu).

© 2014 L. C. Auchincloss *et al.* CBE—Life Sciences Education © 2014 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.

(e.g., American Association for the Advancement of Science [AAAS], 2011). These calls are based on a growing body of research that documents how students benefit from research experiences (Kremer and Bringle, 1990; Kardash, 2000; Rauckhorst *et al.*, 2001; Hathaway *et al.*, 2002; Bauer and Bennett, 2003; Lopatto, 2004, 2007; Lopatto and Tobias, 2010; Seymour *et al.*, 2004; Hunter *et al.*, 2007; Russell *et al.*, 2007; Laursen *et al.*, 2010; Thiry and Laursen, 2011). Undergraduates who participate in research internships (also called research apprenticeships, undergraduate research experiences, or research experiences for undergraduates [REUs]) report positive outcomes, such as learning to think like a scientist, finding research exciting, and intending to pursue graduate education or careers in science (Kardash, 2000; Laursen *et al.*, 2010; Lopatto and Tobias, 2010). Research experiences are thought to be especially beneficial for women and underrepresented minority students, presumably because they support the development of relationships with more senior scientists and with peers who can offer critical support to students who might

Table 1. Features of CUREs compared with research internships

	CUREs	Research internships
Scale	Many students	Few students
Mentorship structure	One instructor to many students	One instructor to one student
Enrollment	Open to all students in a course	Open to a selected or self-selecting few
Time commitment	Students invest time primarily in class	Students invest time primarily outside class
Setting	Teaching lab	Faculty research lab

otherwise leave the sciences (Gregerman *et al.*, 1998; Barlow and Villarejo, 2004; Eagan *et al.*, 2011). Yet most institutions lack the resources to involve all or even most undergraduates in a research internship (Wood, 2003; Desai *et al.*, 2008; Harrison *et al.*, 2011).

Faculty members have developed alternative approaches to engage students in research with the aim of offering these educational benefits to many more students (Wei and Woodin, 2011). One approach that is garnering increased attention is what we call a *course-based undergraduate research experience*, or CURE. CUREs involve whole classes of students in addressing a research question or problem that is of interest to the scientific community. As such, CUREs have the potential to expand undergraduates' access to and involvement in research. We illustrate this in Table 1 by comparing CUREs with research internships, in which undergraduates work one-on-one with a mentor, either a graduate student, technician, postdoctoral researcher, or faculty member.

CUREs offer the capacity to involve many students in research (e.g., Rowland *et al.*, 2012) and can serve all students who enroll in a course—not only self-selecting students who seek out research internships or who participate in specialized programs, such as honors programs or programs that support research participation by disadvantaged students. Moreover, CUREs can be integrated into introductory-level courses (Dabney-Smith, 2009; Harrison *et al.*, 2011) and thus have the potential to exert a greater influence on students' academic and career paths than research internships that occur late in an undergraduate's academic program and thus serve primarily to confirm prior academic or career choices (Hunter *et al.*, 2007). Entry into CUREs is logistically straightforward; students simply enroll in the course. Research internships often require an application (e.g., to REU sites funded by the National Science Foundation [NSF]) or searching and networking to find faculty interested in involving undergraduates in research. For students, CUREs may reduce the stress associated with balancing a research internship with course work during a regular academic term (Rowland *et al.*, 2012). CUREs may also offer different types of opportunities for students to develop ownership of projects, as they ask their own questions or analyze their own samples. Although this can be the case for research internships, it may be less common, given the pressure on research groups to complete and publish the work outlined in grant proposals. In both environments, beginning undergraduate researchers more often contribute to ongoing projects rather than developing their own independent projects. Opportunities for the latter are important, as work from Hanauer and colleagues (2012) suggests that students' development

of a sense of ownership can contribute to their persistence in science.

The Course-Based Undergraduate Research Experiences Network (CUREnet; <http://curenet.franklin.uga.edu>) was initiated in 2012 with funding from NSF to support CURE instruction by addressing topics, problems, and opportunities inherent to integrating research experiences into undergraduate courses. During early discussions, the CUREnet community identified a need for a clearer definition of what constitutes a CURE and a need for systematic exploration of how students are affected by participating in CUREs. Thus, a small working group with expertise in CURE design and assessment was assembled in September 2013 to:

1. Draft an operational definition of a CURE;
2. Summarize research on CUREs, as well as findings from studies of undergraduate research internships that would be useful for thinking about how students are influenced by participating in CUREs; and
3. Identify areas of greatest need with respect to evaluation of CUREs and assessment of CURE outcomes.

In this paper, we summarize the meeting discussion and offer recommendations for next steps in the assessment of CUREs.

CUREs DEFINED

The first aim of the meeting was to define a CURE. We sought to answer the question: How can a CURE be distinguished from other laboratory learning experiences? This allows us to make explicit to students how a CURE may differ from their other science course work and to distinguish a CURE from other types of learning experiences for the purposes of education research and evaluation. We began by discussing what we mean by "research." We propose that CUREs involve students in the following:

1. Use of scientific practices. Numerous policy documents, as well as an abundance of research on the nature and practice of science, indicate that science research involves the following activities: asking questions, building and evaluating models, proposing hypotheses, designing studies, selecting methods, using the tools of science, gathering and analyzing data, identifying meaningful variation, navigating the messiness of real-world data, developing and critiquing interpretations and arguments, and communicating findings (National Research Council [NRC], 1996; Singer *et al.*, 2006; Duschl *et al.*, 2007; Bruck *et al.*, 2008; AAAS, 2011; Quinn *et al.*,

2011). Individuals engaged in science make use of a variety of techniques, such as visualization, computation, modeling, and statistical analysis, with the aim of generating new scientific knowledge and understanding (Duschl *et al.*, 2007; AAAS, 2011). Although it is unrealistic to expect students to meaningfully participate in all of these practices during a single CURE, we propose that the opportunity to engage in multiple scientific practices (e.g., not only data collection) is a CURE hallmark.

2. **Discovery.** Discovery is the process by which new knowledge or insights are obtained. Science research aims to generate new understanding of the natural world. As such, discovery in the context of a CURE implies that the outcome of an investigation is unknown to both the students and the instructor. When the outcomes of their work are not predetermined, students must make decisions such as how to interpret their data, when to track down an anomaly and when to ignore it as “noise,” or when results are sufficiently convincing to draw conclusions (Duschl *et al.*, 2007; Quinn *et al.*, 2011). Discovery carries with it the risk of unanticipated outcomes and ambiguous results because the work has not been done before. Discovery also necessitates exploration and evidence-based reasoning. Students and instructors must have some familiarity with the current body of knowledge in order to contribute to it and must determine whether the new evidence gathered is sufficient to support the assertion that new knowledge has been generated (Quinn *et al.*, 2011). We propose that discovery in the context of a CURE means that students are addressing novel scientific questions aimed at generating and testing new hypotheses. In addition, when their work is considered collectively, students’ findings offer some new insight into how the natural world works.
3. **Broadly relevant or important work.** Because CUREs provide opportunities for students to build on and contribute to current science knowledge, they also present opportunities for impact and action beyond the classroom. In some CUREs, this may manifest as authorship or acknowledgment in a science research publication (e.g., Leung *et al.*, 2010; Pope *et al.*, 2011). In other CUREs, students may develop reports of interest to the local community, such as a report on local water quality or evidence-based recommendations for community action (e.g., Savan and Sider, 2003). We propose that CUREs involve students in work that fits into a broader scientific endeavor that has meaning beyond the particular course context. (We choose the language of “broader relevance or importance” rather than the term “authenticity” because views on the authenticity of a learning experience may shift over time [Rahm *et al.*, 2003] and may differ among students, instructors, and the broader scientific community.)
4. **Collaboration.** Science research increasingly involves teams of scientists who contribute diverse skills to tackling large and complex problems (Quinn *et al.*, 2011). We propose that group work is not only a common practical necessity but also an important pedagogical element of CUREs because it exposes students to the benefits of bringing together many minds and hands to tackle a problem (Singer *et al.*, 2006). Through collaboration,

students can improve their work in response to peer feedback. Collaboration also develops important intellectual and communication skills as students verbalize their thinking and practice communicating biological ideas and interpretations either to fellow students in the same discipline or to students in other disciplines. This may also encourage students’ metacognition—solidifying their thinking and helping them to recognize shortcomings in their knowledge and reasoning (Chi *et al.*, 1994; Lyman, 1996; Smith *et al.*, 2009; Tanner, 2009).

5. **Iteration.** Science research is inherently iterative because new knowledge builds on existing knowledge. Hypotheses are tested and theories are developed through the accumulation of evidence over time by repeating studies and by addressing research questions using multiple approaches with diverse methods. CUREs generally involve students in iterative work, which can occur at multiple levels. Students may design, conduct, and interpret an investigation and, based on their results, repeat or revise aspects of their work to address problems or inconsistencies, rule out alternative explanations, or gather additional data to support assertions (NRC, 1996; Quinn *et al.*, 2011). 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. Students learn by trying, failing, and trying again, and by critiquing one another’s work, especially the extent to which claims can be supported by evidence (NRC, 1996; Duschl *et al.*, 2007; Quinn *et al.*, 2011).

These activities, when considered in isolation, are not unique to CUREs. Rather, we propose that it is the integration of all five dimensions that makes a learning experience a CURE. Of course, CUREs will vary in the frequency and intensity of each type of activity. We present the dimensions in Table 2 and delineate how they are useful for distinguishing between the following four laboratory learning environments:

1. A traditional laboratory course, in which the topic and methods are instructor defined; there are clear “cook-book” directions and a predetermined outcome that is known to students and to the instructor (Domin, 1999; Weaver *et al.*, 2008);
2. An inquiry laboratory course, in which students participate in many of the cognitive and behavioral practices that are commonly performed by scientists; typically, the outcome is unknown to students, and they may be challenged to generate their own methods. The motivation for the inquiry is to challenge the students, rather than contribute to a larger body of knowledge (Domin 1999; Olson and Loucks-Horsley, 2000; Weaver *et al.*, 2008);
3. A CURE, in which students address a research question or problem that is of interest to the broader community with an outcome that is unknown both to the students and to the instructor (Domin 1999; Bruck *et al.*, 2008; Weaver *et al.*, 2008); and

Table 2. Dimensions of different laboratory learning contexts

Dimension		Traditional	Inquiry	CURE	Internship
Use of science practices	Students engage in ...	Few scientific practices	Multiple scientific practices	Multiple scientific practices	Multiple scientific practices
	Study design and methods are ...	Instructor driven	Student driven	Student or instructor driven	Student or instructor driven
Discovery	Purpose of the investigation is ...	Instructor defined	Student defined	Student or instructor defined	Student or instructor defined
	Outcome is ...	Known to students and instructors	Varied	Unknown	Unknown
Broader relevance or importance	Findings are ...	Previously established	May be novel	Novel	Novel
	Relevance of students' work ...	Is limited to the course	Is limited to the course	Extends beyond the course	Extends beyond the course
	Students' work presents opportunities for action ...	Rarely	Rarely	Often	Often
Collaboration	Collaboration occurs ...	Among students in a course	Among students in a course	Among students, teaching assistants, instructor in a course	Between student and mentor in a research group
	Instructor's role is ...	Instruction	Facilitation	Guidance and mentorship	Guidance and mentorship
Iteration	Risk of generating "messy" data are ...	Minimized	Significant	Inherent	Inherent
	Iteration is built into the process ...	Not typically	Occasionally	Often	Often

4. A research internship, in which a student is apprenticed to a senior researcher (faculty, postdoc, grad student, etc.) to help advance a science research project (Seymour *et al.*, 2004).

The five dimensions comprise a framework that can be tested empirically by characterizing how a particular dimension is manifested in a program, developing scales to measure the degree or intensity of each dimension, and determining whether the dimensions in part or as a whole are useful for distinguishing CUREs from other laboratory learning experiences. Once tested, we believe that this framework will be useful to instructors, institutional stakeholders, education researchers, and evaluators.

Instructors may use the framework to delineate their instructional approach, clarify what students will be expected to do, and articulate their learning objectives. For example, in traditional laboratory instruction, students may collect and analyze data but generally do not build or evaluate models or communicate their findings to anyone except the instructor. During inquiry laboratory instruction, students may be able to complete a full inquiry cycle and thus engage at some level in the full range of scientific practices. Students in CUREs and research internships may engage in some scientific practices in depth, but neglect others, depending on the particular demands of the research and the structure of the project. As instructors define how their course activities connect to desired student outcomes, they can also identify directions for formative and summative assessment.

Education researchers and evaluators may use the framework to characterize particular instructional interventions with the aim of determining which dimensions, to what degree and intensity, correlate with desired student outcomes. For instance, students who engage in the full range of scientific practices could reasonably be expected to improve their skills across the range of practices, while students who participate in only a subset of practices can only be expected to improve in those specific practices. Similarly, the extent to which students have control over the methods they employ may influence their sense of ownership over the investigation, thus increasing their motivation and perhaps contributing to their self-identification as scientists. Using this framework to identify critical elements of CUREs and how they relate (or not) to important student outcomes can inform both the design of CUREs and their placement in a curriculum.

CURRENT KNOWLEDGE FROM ASSESSMENT OF CUREs

With this definition in mind, the meeting then turned to summarizing what is known from the study of CUREs, primarily in biology and chemistry. Assessment and evaluation of CUREs has been limited to a handful of multisite programs (e.g., Goodner *et al.*, 2003; Hatfull *et al.*, 2006; Lopatto *et al.*, 2008, Caruso *et al.*, 2009; Shaffer *et al.*, 2010; Harrison *et al.*, 2011) and projects led by individual instructors (e.g., Drew and Triplett 2008; Siritunga *et al.*, 2011). For the most part, these studies have emphasized student perceptions of the

outcomes they realize from participating in course-based research, such as the gains they have made in research skills or clarification of their intentions to pursue further education or careers in science. To date, very few studies of student learning during CUREs have been framed according to learning theories. With a few exceptions, studies of CUREs have not described pathways that students take to arrive at specific outcomes—in other words, what aspects of the CURE are important for students to achieve both short- and long-term gains.

Some studies have compared CURE instruction with research internships and have found, in general, that students report many of the same gains (e.g., Shaffer *et al.*, 2010). A handful of studies have compared student outcomes from CUREs with those from other laboratory learning experiences. For example, Russell and Weaver (2011) compared students' views of the nature of science after completing a traditional laboratory, an inquiry laboratory, or a CURE. The researchers used an established approach developed by Lederman and colleagues (2002) to assess students' views of the nature of science, but it is not clear whether students in this study chose to enroll in a traditional or CURE course or whether the groups differed in other ways that might influence the extent to which their views changed following their lab experiences. Students in all three environments—traditional, inquiry, and CURE—made gains in their views of the nature of scientific knowledge as experimental and theory based, but only students in the CURE showed progress in their views of science as creative and process based. When students who participated in a CURE or a traditional lab were queried 2 or 3 yr afterward, they continued to differ in their perceptions of the gains they made in understanding how to do research and in their confidence in doing research (Sztejnberg and Weaver, 2013).

In another study, Rowland and colleagues (2012) compared student reports of outcomes from what they called an *active-learning laboratory undergraduate research experience* (ALLURE, which is similar to a CURE) with those from a traditional lab course. Students could choose the ALLURE or traditional instruction, which may have resulted in a self-selection bias. Students in both environments reported increased confidence in their lab skills, including technical skills (e.g., pipetting) and analytical skills (e.g., deciding whether one experimental approach is better than another). Generally, students reported similar skill gains in both environments, indicating that students can develop confidence in their lab skills during both traditional and CURE/ALLURE experiences.

Most studies reporting assessment of CUREs in the life sciences have made use of the Classroom Undergraduate Research Experiences (CURE) Survey (Lopatto and Tobias, 2010). The CURE Survey comprises three elements: 1) instructor report of the extent to which the learning experience resembles the practice of science research (e.g., the outcomes of the research are unknown, students have some input into the focus or design of the research); 2) student report of learning gains; and 3) student report of attitudes toward science. A series of Likert-type items probe students' attitudes toward science and their educational and career interests, as well as students' perceptions of the learning experience, the nature of science, their own learning styles, and the science-related skills they developed from participating in a CURE. Use of the CURE Survey has been an important first step in assessing

student outcomes of these kinds of experiences. Yet this instrument is limited as a measure of the nature and outcomes of CUREs because some important information is missing about its overall validity. No information is available about its dimensionality—that is do student responses to survey items meant to represent similar underlying concepts correlate with each other, while correlating less with items meant to represent dissimilar concepts? For example, do responses to items about career interests correlate with themselves highly, but correlate less with items focused on attitudes toward science, a dissimilar concept? Other validity questions are also not addressed. For instance, does the survey measure all important aspects of CUREs and CURE outcomes, or are important variables missing? Is the survey useful for measuring a variety of CUREs in different settings, such as CUREs for majors or nonmajors, or CUREs at an introductory or advanced levels? Finally, is the survey a reliable measure—does the survey measure outcomes consistently over time and across different individuals and settings? To be consistent with the definition of CUREs given above, an assessment instrument must both touch on all five dimensions and elicit responses that capture other important aspects of CURE instruction that may be missing from this description. This will help ensure that the instrument has “content validity” (Trochim, 2006), meaning that the instrument can be used to measure all of the features important in a CURE learning experience.

The CURE Survey relies on student perceptions of their own knowledge and skill gains, and like other such instruments, it is subject to concerns about the validity of self-report of learning gains. There is a very broad range of correlations between self-report measures of learning and measurements such as tests or expert judgments. Depending on which measures are compared, there may be a strong correlation, or almost no correlation, between self-reported data and relevant criteria (Falchikov and Boud, 1989). Validity problems with self-assessment can result from poor survey design, with survey items interpreted differently by different students, or from items designed in such a way that students are unable to recall key information or experiences (Bowman 2011; Porter *et al.*, 2011). The tendency of respondents to give socially desirable answers is a familiar problem with self-reporting. Bowman and Hill (2011) found that student self-reporting of educational outcomes is subject to social bias; students respond more positively because they are either implicitly or explicitly aware of the desired response. A guarantee of anonymity mitigates this validity threat (Albanese *et al.*, 2006). Respondents also give more valid responses when they have a clear idea of what they are assessing and have received frequent and clear feedback about their progress and abilities from others, and when respondents can remember what they did during the assessment period (Kuh, 2001). For example, in her study of the outcomes of undergraduate science research internships, Kardash (2000) compared perceptions of both student interns and faculty mentors of the gains interns made from participating in research. She found good agreement between interns and mentors on some skills, such as understanding concepts in the field and collecting data, but statistically significant differences between mentor and intern ratings of other skills, with interns rating themselves more positively on their understanding of the importance of controls in research, their abilities to interpret results in light of original hypotheses, and their abilities to relate results to

the “bigger picture.” More research is needed to understand the extent to which different students (majors, nonmajors, introductory, advanced, etc.) are able to accurately self-assess the diverse knowledge and skills they may develop from participating in CUREs.

A few studies have focused on the psychosocial outcomes of participating in CUREs. One such study, conducted by Hanauer and colleagues (2012), documented the extent to which students developed a sense of ownership of the science projects they completed in a traditional laboratory course, a CURE involving fieldwork, or a research internship. Using linguistic analysis, the authors found that students in the CURE reported a stronger sense of ownership of their research projects compared with students who participated in traditional lab courses and research internships (Hanauer *et al.*, 2012; Hanauer and Dolan, *in press*, 2014); these students also reported higher levels of persistence in science or medicine (Hanauer *et al.*, 2012). Although the inferred relationship needs to be explored with a larger group of students and a more diverse set of CUREs, these results suggest that it is important to consider ownership and other psychosocial outcomes in future research and evaluation of CUREs.

A few studies have explored whether and how different students experience CUREs differently and, in turn, realize different outcomes from CUREs. This is an especially noteworthy gap in the knowledge base, given the calls to engage all students in research experiences and that research has suggested that different students may realize different outcomes from participating in research (e.g., AAAS, 2011; Thiry *et al.*, 2012). In one such study, Alkahrer and Dolan (*in press*, 2014) interviewed students enrolled in a CURE, the Partnership for Research and Education in Plants for Undergraduates, at three different types of institutions (i.e., community college, liberal arts college, research university) in order to examine whether and how their sense of scientific self-authorship shifted during the CURE. Baxter-Magolda (1992) defined self-authorship as the “internal capacity to define one’s beliefs, relations, and social identity” or, in this context, how one sees oneself with respect to science knowledge—as a consumer, user, or producer. Developing a sense of scientific self-authorship may be an important predictor of persistence in science, as students move from simply consuming science knowledge as it is presented to becoming critical users of science, and to seeing themselves as capable of contributing to the scientific body of knowledge. Alkahrer and Dolan (*in press*, 2014) found that some CURE students made progress in their self-authorship because they perceived the CURE goals as important to the scientific community, yet the tasks were within their capacity to make a meaningful contribution. In contrast, other students struggled with the discovery nature of the CURE in comparison with their prior traditional lab learning experiences. They perceived their inability to find the “right answer” as reflecting their inability to do science. More research is needed to determine whether and how students’ backgrounds, motives, and interests influence how they experience CUREs, and whether they realize different outcomes as a result.

NEXT STEPS FOR CURE ASSESSMENT

Our discussion and collective knowledge of research on CUREs and undergraduate research internships revealed

several gaps in our understanding of CUREs, which can be addressed by:

- Defining frameworks and learning theories that may help explain how students are influenced by participating in CUREs, and utilizing these frameworks or theories to design and study CUREs;
- Identifying and measuring the full range of important outcomes likely to occur in CURE contexts;
- Using valid and reliable measures, some of which have been used to study research internships or other undergraduate learning experiences and could be adapted for CURE use, as well as developing and testing new tools to assess CUREs specifically (see Weiss and Sosulski [2003] or Trochim [2006] for general explanations of validity and reliability in social science measurement);
- Establishing which outcomes are best documented using self-reporting, and developing new tools or adapting existing tools to measure other outcomes; and
- Gathering empirical evidence to identify the distinctive dimensions of CUREs and ways to characterize the degree to which they are present in a given CURE, as well as conducting investigations to characterize relationships between particular CURE dimensions or activities and student outcomes.

Following these recommendations will require a collective, scholarly effort involving many education researchers and evaluators and many CUREs that are diverse in terms of students, instructors, activities, and institutional contexts. We suggest that priorities of this collective effort should be to:

1. Use current knowledge from the study of CUREs, research internships, and other relevant forms of laboratory instruction (e.g., inquiry) to define short-, medium-, and long-term outcomes that may result from student participation in CUREs;
2. Observe and characterize many diverse CUREs to identify the activities within CUREs likely to directly result in these short-term outcomes, delineating both rewards and difficulties students encounter as they participate;
3. Use frameworks or theories and current knowledge to hypothesize pathways students may take toward achieving long-term outcomes—the connections between activities and short-, medium-, and long-term outcomes;
4. Determine whether one can identify key short- and medium-term outcomes that serve as important “linchpins” or connecting points through which students progress to achieve desired long-term outcomes; and
5. Assess the extent to which students achieve these key outcomes as a result of CURE instruction, using existing or novel instruments (e.g., surveys, interview protocols, tests) that have been demonstrated to be valid and reliable measures of the desired outcomes.

At the front end, this process will require increased application of learning theories and consideration of the supporting research literature, but it is likely to result in many highly testable hypotheses and a more focused and informative approach to CURE assessment overall. For example, if we can define pathways from activities to outcomes, instructors will

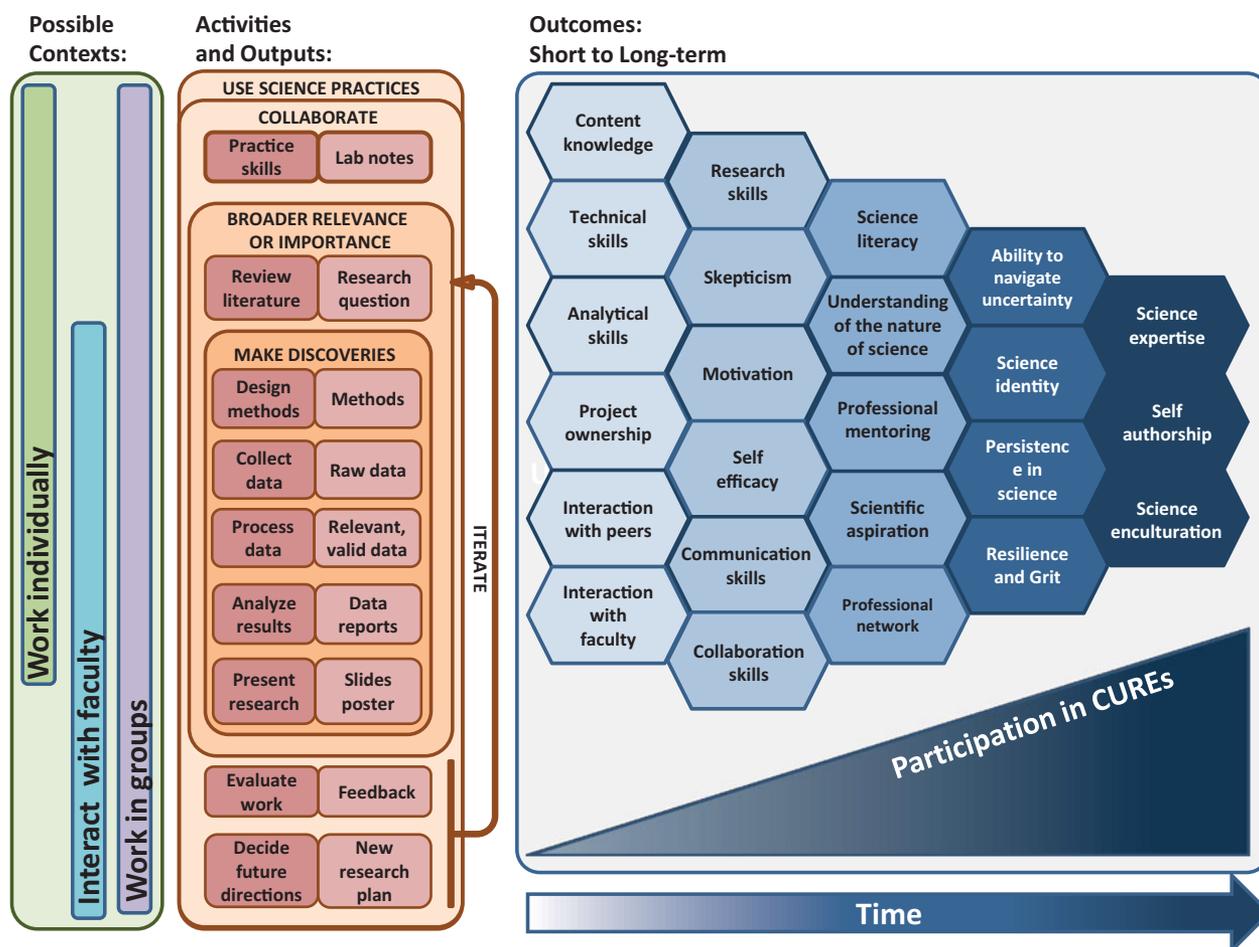


Figure 1. CURE logic model. This model depicts the set of variables at play in CUREs identified by the authors. During CUREs, students can work individually, in groups, or with faculty (context, green box on left) to perform corresponding activities (middle, red boxes) that yield measurable outputs (middle, pink boxes). Activities and outputs are grouped according to the five related elements of CUREs (orange boxes and arrow). Possible CURE outcomes (blue) are ordered left to right according to when students might be able to demonstrate the outcome (blue arrow) and whether the outcome is likely to be achievable from participation in a single vs. multiple CUREs (blue triangle).

be better able to select activities to include or emphasize during CURE instruction and decide which short-term outcomes to assess. Education researchers and evaluators will be better able to hypothesize which aspects of CURE instruction are most critical for desired student outcomes and the most salient to study.

Drawing from many of the references cited in this report, we have drafted a logic model for CURE instruction (Figure 1) as the first step in this process. (For more on logic models, see guidance from the W. K. Kellogg Foundation [2006].) The model includes the range of contexts, activities, outputs, and outcomes of CUREs that arose during our discussion. The model also illustrates hypothetical relationships between time, participation in CUREs, and short- and long-term outcomes resulting from CURE activities.

It is important to recognize that, given the limited time frame and scope of any single CURE, students will not participate in all possible activities or achieve all possible outcomes depicted in the model. Rather, CURE instructors or evaluators could define a particular path and use it as a guide for designing program evaluations and assessing student out-

comes. Figure 2 presents an example of how to do this with a focus on a subset of CURE activities and outcomes. It is a simplified pathway model based on findings from the research on undergraduate research internships and CUREs summarized above. Boxes in this model are potentially measurable waypoints, or steps, on a path that connects student participation in three CURE activities with the short-term outcomes students may realize during the CURE, medium-term outcomes they may realize at the end of or after the CURE, and potential long-term outcomes. Although each pathway is supported by evidence or hypotheses from the study of CUREs and research internships, these are not the only means to achieve long-term outcomes, and they do not often act alone. Rather, the model is intended to illustrate that certain short- and medium-term outcomes are likely to have a positive effect on linked long-term outcomes. See Urban and Trochim (2009) for a more detailed discussion of this approach.

We explain below the example depicted in Figure 2, referencing explicit waypoints on the path with italics. This model is grounded in situated-learning theory (Lave and Wenger, 1991), which proposes that learning involves engagement in

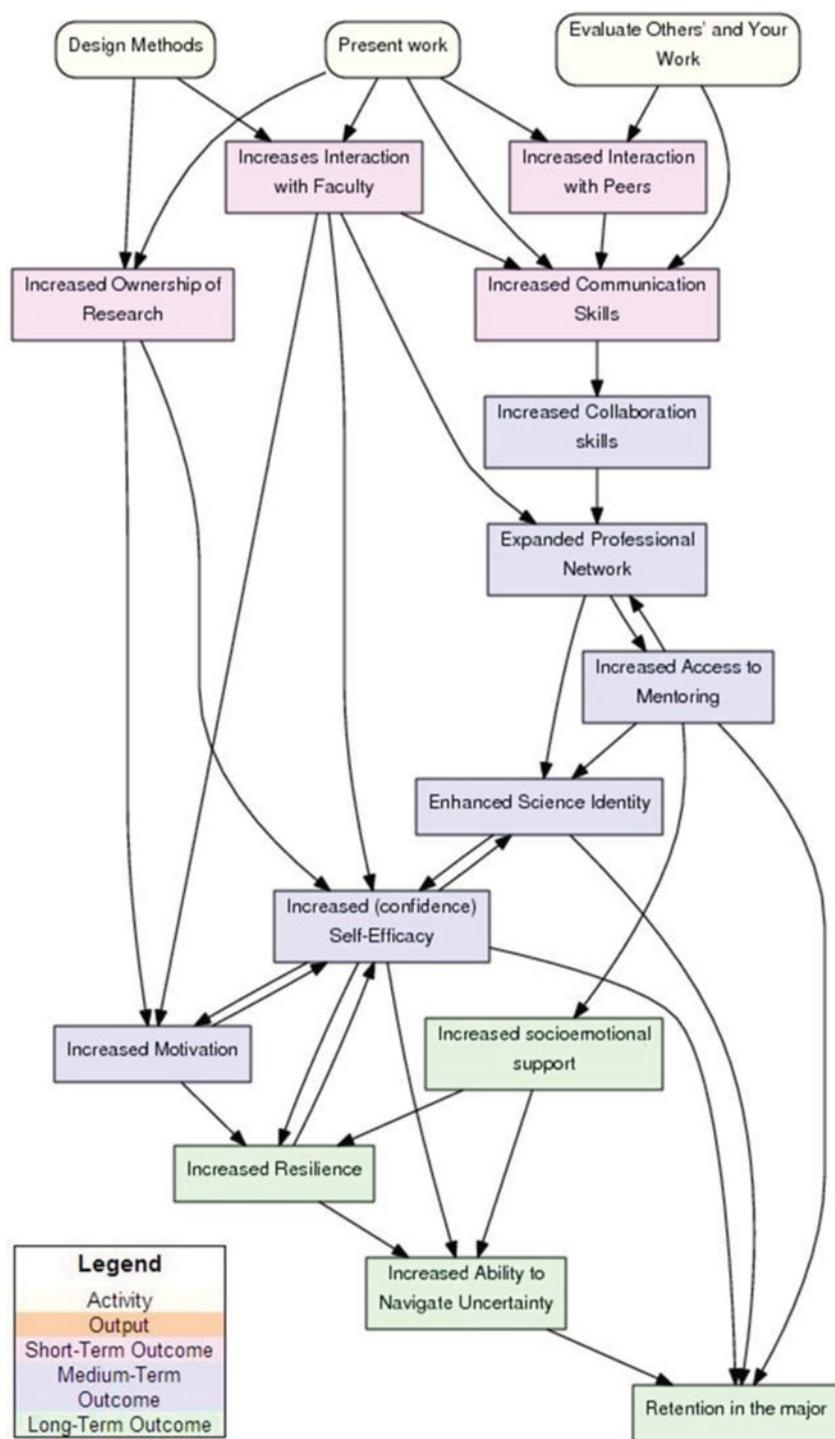


Figure 2. Example of a pathway model to guide CURE assessment. This model identifies a subset of activities (beige) students are likely to do during a CURE and the short- (pink), medium- (blue), and long- (green) term outcomes they may experience as a result. The arrows depict demonstrated or hypothesized relationships between activities and outcomes. (This figure is generated using software from the Cornell Office of Research and Evaluation [2010].)

a “community of practice,” a group of people working on a common problem or endeavor (e.g., addressing a particular research question) and using a common set of practices (e.g., science practices). Situated-learning theory envisions learning as doing (e.g., presenting and evaluating work) and as belonging (e.g., interacting with faculty and peers, building networks), factors integral to becoming a practitioner (Wenger, 2008)—in the case of CUREs, becoming a scientist.

Retention in a science major is a desired and measurable long-term outcome (bottom of Figure 2) that indicates students are making progress in becoming scientists and has been shown to result from participation in research (Perna *et al.*, 2009; Eagan *et al.*, 2013). Based on situated-learning theory, we hypothesize that three activities students might engage in are likely to lead to *retention in a science major*: *design methods*, *present their work*, and *evaluate their own and others’ work*

during their research experience (Caruso *et al.*, 2009; Harrison *et al.*, 2011; Hanauer *et al.*, 2012). These activities reflect the dimensions of “use of scientific practices” and “collaboration” described above.

Following the right-hand path in the model, when students *present their work* and *evaluate their own and others’ work*, they will likely interact with each other and with faculty (Eagan *et al.*, 2011). *Interactions with faculty* and *interactions with peers* may lead to improvements in students’ *communication* and *collaboration skills*, including their abilities to defend their work, negotiate, and make decisions about their research based on interactions (Ryder *et al.*, 1999; Alexander *et al.*, 2000; Seymour *et al.*, 2004). Through these interactions, students may expand their *professional networks*, which may in turn offer increased *access to mentoring* (Packard, 2004; Eagan *et al.*, 2011). Mentoring relationships, especially with faculty, connect undergraduates to networks that promote their education and career development by building their sense of scientific identity and defining their role within the broader scientific community (Crisp and Cruz, 2009; Hanauer, 2010; Thiry *et al.*, 2010; Thiry and Laursen, 2011; Stanton-Salazar, 2011). Peer and faculty relationships also offer socio-emotional support that can foster students’ resilience and their ability to navigate the uncertainty inherent to science research (Chemers *et al.*, 2011; Thiry and Laursen, 2011). Finally, research on factors that lead to retention in science majors indicates that increased *science identity* (Laursen *et al.*, 2010; Estrada *et al.*, 2011), *ability to navigate uncertainty*, and *resilience* are important precursors to a sense of belonging and ultimate retention (Gregerman *et al.*, 1998; Zeldin and Pajares, 2000; Maton and Hrabowski, 2004; Seymour *et al.*, 2004). The model also suggests that *access to mentoring* is a linchpin, a short- to medium-term outcome that serves as a connecting point through which activities are linked to long-term outcomes. Thus, access to mentoring might be assessed to diagnose students’ progress along the top pathway and predict the likelihood that they will achieve long-term outcomes. (For more insight into why assessing linchpins is particularly informative, see Urban and Trochim [2009].)

Examples of measures that may be useful for testing aspects of this model and for which validity and reliability information is available include: the scientific identity scale developed by Chemers and colleagues (2011) and revised by Estrada and colleagues (2011); the student cohesiveness, teacher support, and cooperation scales of the What Is Happening in This Class? questionnaire (Dorman, 2003); and the faculty mentorship items published by Eagan and colleagues (2011). Data will need to be collected and analyzed using standard validation procedures to determine the usefulness of these scales for studying CUREs. Qualitative data from interviews or focus groups can be used to determine that students perceive these items as measuring relevant aspects of their CURE experiences and to confirm that they are interpreting the questions as intended. For example, developers of the Undergraduate Research Student Self-Assessment instrument used extensive interview data to identify key dimensions of student outcomes from research apprenticeship experiences, and then think-aloud interviews to test and refine the wording of survey items (Hunter *et al.*, 2009). Interviews can also establish whether items apply to different groups of students. For example, items in the scientific identity scale (e.g., “I feel like I belong in the field of science”)

may seem relevant, and thus “valid,” to science majors but not to non-science majors. Similarly, the faculty-mentoring items noted above (Eagan *et al.*, 2011) include questions about whether faculty provided, for example, “encouragement to pursue graduate or professional study” or “an opportunity to work on a research project.” The first item will be most relevant to students who are enrolled in an advanced rather than an introductory CURE, while the second may be relevant only to students early enough in their undergraduate careers to have time to pursue a research internship. In addition, students may interpret the phrase “opportunity to work on a research project” in ways that are unrelated to mentorship by faculty, especially in the context of a CURE class with its research focus. Statistical analyses (e.g., factor analysis, calculation of Cronbach’s alpha; Netemeyer *et al.*, 2003) should confirm that the scales are consistent and stable—are they measuring what they are intended to measure and do they do so consistently? Such analyses would help determine whether students are responding as anticipated to particular items or scales and whether instruments developed to measure student outcomes of research internships can detect student growth from participation in CUREs, which are different experiences.

We can also follow the left-hand path in this model with a focus on the CURE activities of *designing methods* and *presenting work*. This path is grounded in Baxter Magolda’s (2003) work on students’ epistemological development and her theory of self-authorship. Specifically, as students take ownership of their learning, they transition from seeing themselves as consumers of knowledge to seeing themselves as producers of knowledge. Some students who design their own methods and present their work report an increased sense of *ownership of the research* (Hanauer *et al.*, 2012; Hanauer and Dolan, 2014). Increased ownership has been shown to improve *motivation* and *self-efficacy*. *Self-efficacy* and *motivation* work in a positive-feedback loop to enhance one another and contribute to development of long-term outcomes, such as increased *resilience* (Graham *et al.*, 2013). Social cognitive theory is useful for explaining this relationship: if people believe they are capable of accomplishing a task—described in the literature as *self-efficacy*—they are more likely to put forth effort, *persist* in the task, and be *resilient* in the face of failure (Bandura, 1986; Zeldin and Pajares, 2000). *Self-efficacy* has also been positively related to *science identity* (Zeldin and Pajares, 2000; Seymour *et al.*, 2004; Hanauer, 2010; Estrada *et al.*, 2011; Adedokun *et al.*, 2013). Thus, *self-efficacy* becomes a linchpin that interacts closely with *motivation* and can be connected to retention in a science major. Existing measures that may be useful for testing this model and for which validity and reliability information is available include: the Project Ownership Survey (Hanauer and Dolan, 2014), scientific self-efficacy and scientific identity scales (Chemers *et al.*, 2011; Estrada *et al.*, 2011); and the self-authorship items from the Career Decision Making Survey (Creamer *et al.*, 2010). Again, data would need to be collected and analyzed using standard validation procedures to determine the usefulness of these scales for studying CUREs.

When considering what to include in a model or which pathways to emphasize, we encourage CURE stakeholders to remember that each CURE is in its own stage of development and has its own life cycle. Some are just starting and others are well established. CUREs at the beginning stages of

implementation are likely to be better served by evaluating how well the program is being implemented before evaluating downstream student outcomes. Thus, early in the development of a CURE, those who are assessing CUREs may want to model a limited set of activities, outputs, and short-term outcomes. CUREs at later stages of development may focus more of their evaluation efforts on long-term student outcomes because earlier evaluations have demonstrated stability of the program's implementation. At this point, findings regarding student outcomes can more readily be attributed to participation in the CURE.

Last, we would like to draw some comparisons between CUREs and research internships because these different experiences are likely to offer unique and complementary ways of engaging undergraduates in research that could be informative for CURE assessment. As noted above, a handful of studies indicate that CURE students may realize some of the same outcomes observed for students in research internships (Goodner *et al.*, 2003; Drew and Triplett 2008; Lopatto *et al.*, 2008; Caruso *et al.*, 2009; Shaffer *et al.*, 2010; Harrison *et al.*, 2011). Yet, differences between CUREs and research internships (Table 1) are likely to influence the extent to which students achieve any particular outcome. For example, CUREs may offer different opportunities for student input and autonomy (Patel *et al.*, 2009; Hanauer *et al.*, 2012; Hanauer and Dolan, 2014; Table 2). The structure of CUREs may allow undergraduates to assume more responsibility in project decision making and take on leadership roles that are less often available in research internships. CUREs may involve more structured group work, providing avenues for students to develop analytical and collaboration skills as they explain or defend their thinking and provide feedback to one another. In addition, CURE students may have increased opportunities to develop and express skepticism because they are less likely to see their peers as authority figures.

Alternatively, some CURE characteristics may limit the nature or extent of outcomes that students realize. CUREs take place in classroom environments with a much higher student-faculty ratio than is typical of UREs. With fewer experienced researchers to model scientific practices and provide feedback, students may be less likely to develop a strong understanding of the nature of science or a scientific identity. The amount of time students may spend doing the work in a CURE course is likely to be significantly less than what they would spend in a research internship. Students who enroll in CURE courses may be less interested in research, which may affect their own and classmates' motivation and longer-term outcomes related to motivation. Research interns are more likely to develop close collegial relationships with faculty and other researchers, such as graduate students, postdoctoral researchers, and other research staff, who can in turn expand their professional network. In addition, CURE instructors may have limited specialized knowledge of the science that underpins the CURE. Thus, CURE students may not have access to sufficient mentorship or expertise to maximize the scientific and learning outcomes.

SUMMARY

This report is a first attempt to capture the distinct characteristics of CUREs and discuss ways in which they can be

systematically evaluated. Utilizing current research on CUREs and on research internships, we identify and describe five dimensions of CURE instruction: use of science practices, discovery, broader relevance or importance, iteration, and collaboration. We describe how these elements might vary among different laboratory learning experiences and recommend an approach to CURE assessment that can characterize CURE activities and outcomes. We hope that our discussion draws attention to the importance of developing, observing, and characterizing many diverse CUREs. We also hope that this report successfully highlights the enormous potential of CUREs, not only to support students in becoming scientists, but also to provide research experiences to increasing numbers of students who will enter the workforce as teachers, employers, entrepreneurs, and young professionals. We intend for this report to serve as a starting point for a series of informed discussions and education research projects that will lead to far greater understanding of the usages, value, and impacts of CUREs, ultimately resulting in cost-effective, widely accessible, quality research experiences for a large number of undergraduate students.

ACKNOWLEDGMENTS

Thanks to Sandra Laursen for taking a leadership role in organizing this meeting. Support for the meeting was provided by a grant from the NSF (DBI-1061874). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NSF.

REFERENCES

- Adedokun OA, Bessenbacher AB, Parker LC, Kirkham LL, Burgess WD (2013). Research skills and STEM undergraduate research students' aspirations for research careers: Mediating effects of research self efficacy. *J Res Sci Teach* 50, 940–951.
- Albanese M, Dotti S, Mejicano G, Zakowski L, Seibert C, Van Eyck S (2006). Distorted perceptions of competence and incompetence are more than regression effects. *Adv Health Sci Educ* 11, 267–278.
- Alexander BB, Foertsch J, Daffinrud S, Tapia R (2000). The Spend a Summer with a Scientist (SaS) Program at Rice University: a study of program outcomes and essential elements, 1991–1997. *Counc Undergrad Res Q* 20, 127–133.
- Alkaber I, Dolan EL (2014). Integrating research into undergraduate courses: current practices and future directions. In: *Research in Science Education: Research Based Undergraduate Science Teaching*, ed. D Sunal, C Sunal, D Zollman, C Mason, and E Wright (*in press*).
- American Association for the Advancement of Science (2011). *Vision and Change in Undergraduate Biology Education*, Washington, DC.
- Bandura A (1986). The explanatory and predictive scope of self-efficacy theory. *J Clin Soc Psychol* 4, 359–373.
- Barlow AE, Villarejo M (2004). Making a difference for minorities: evaluation of an educational enrichment program. *J Res Sci Teach* 41, 861–881.
- Bauer KW, Bennett JS (2003). Alumni perceptions used to assess undergraduate research experience. *J High Educ* 74, 210–230.
- Baxter-Magolda MB (1992). *Knowing and Reasoning in College: Gender-related Patterns in Students' Intellectual Development*, San Francisco: Jossey-Bass.
- Baxter Magolda MB (2003). Identity and learning: student affairs' role in transforming higher education. *J Coll Student Dev* 44, 231–247.

- Bowman NA (2011). Examining systematic errors in predictors of college student self-reported gains. *New Direct Institut Res* 150, 7–19.
- Bowman NA, Hill PL (2011). Measuring how college affects students: social desirability and other potential biases in college student self reported gains. *New Direct Institut Res* 150, 73–85.
- Bruck L, Bretz SL, Towns ML (2008). Characterizing the level of inquiry in the undergraduate laboratory. *J Coll Sci Teach* 38, 52–88.
- Caruso SM, Sandoz J, Kelsey J (2009). Non-STEM undergraduates become enthusiastic phage-hunters. *CBE Life Sci Educ* 8, 278–282.
- Chemers MM, Zurbriggen EL, Syed M, Goza BK, Bearman S (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *J Soc Issues* 67, 469–491.
- Chi MTH, de Leeuw N, Chiu MH, LaVanher C (1994). Eliciting self-explanations improves understanding. *Cogn Sci* 18, 439–477.
- Cornell Office of Research and Evaluation (2010). The Netway: Program and Evaluation Planning. <https://core.human.cornell.edu/research/systems/netway.cfm> (accessed 10 November 2013).
- Creamer EG, Baxter Magolda M, Yue J (2010). Preliminary evidence of the reliability and validity of a quantitative measure of self-authorship. *J Coll Student Dev* 51, 550–562.
- Crisp G, Cruz I (2009). Mentoring college students: a critical review of the literature between 1990 and 2007. *Res High Educ* 50, 525–545.
- Dabney-Smith VL (2009). A multi-level case study analysis of campus-based male initiatives, programs, and practices and the impact of participation on the perceptions of first-year African American male community college students in Texas. PhD Thesis, Austin: University of Texas. Available from ProQuest Dissertations and Theses database: UMI no. 3378673.
- Desai KV, Gatson SN, Stiles TW, Stewart RH, Laine GA, Quick CM (2008). Integrating research and education at research extensive universities with research: intensive communities. *Adv Physiol Educ* 32, 136–141.
- Domin D (1999). A review of laboratory instruction styles. *J Chem Educ* 76, 543–547.
- Dorman JP (2003). Cross-national validation of the *What Is Happening in This Class?* (WIHIC) questionnaire using confirmatory factor analysis. *Learn Environ Res* 6, 231–245.
- Drew JC, Triplett EW (2008). Whole genome sequencing in the undergraduate classroom: outcomes and lessons from a pilot course. *J Microbiol Biol Educ* 9, 3.
- Duschl RA, Schweingruber HA, Shouse AW (eds.) (2007). *Taking Science to School: Learning and Teaching Science in Grades K–8*, Washington, DC: National Academies Press.
- Eagan MK, Hurtado S, Chang MJ, Garcia GA, Herrera FA, Garibay JC (2013). Making a difference in science education: the impact of undergraduate research programs. *Am Educ Res J* 50, 463–713.
- Eagan MK, Sharkness J, Hurtado S, Mosqueda CM, Chang MJ (2011). Engaging undergraduates in science research: not just about faculty willingness. *Res High Educ* 52, 151–177.
- Estrada M, Woodcock A, Hernandez PR, Schultz P (2011). Toward a model of social influence that explains minority student integration into the scientific community. *J Educ Psychol* 103, 206–222.
- Falchikov N, Boud D (1989). Student self-assessment in higher education: a meta-analysis. *Rev Educ Res* 59, 395–430.
- Goodner BW, Wheeler CA, Hall PJ, Slater SC (2003). Massively parallel undergraduates for bacterial genome finishing. *ASM News* 69, 12.
- Graham MJ, Frederick J, Byars-Winston A, Hunter AB, Handelsman J (2013). Increasing persistence of college students in STEM. *Science* 341, 1455–1456.
- Gregerman SR, Lerner JS, von Hippel W, Jonides J, Nagda BA (1998). Undergraduate student-faculty research partnerships affect student retention. *Rev High Educ* 22, 55–7.
- Hanauer DI (2010). Laboratory identity: a linguistic landscape analysis of personalized space within a microbiology laboratory. *Critic Inquiry in Language Stud* 7, 152–172.
- Hanauer DI, Dolan EL (2014). The Project Ownership Survey: measuring differences in scientific inquiry experiences. *CBE Life Sci Educ* 13, 149–158.
- Hanauer DI, Frederick J, Fotinakes B, Strobel SA (2012). Linguistic analysis of project ownership for undergraduate research experiences. *CBE Life Sci Educ* 11, 378–385.
- Harrison M, Dunbar D, Ratmanskyy L, Boyd K, Lopatto D (2011). Classroom-based science research at the introductory level: changes in career choices and attitude. *CBE Life Sci Educ* 10, 279–286.
- Hatfull GF *et al.* (2006). Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLoS Genet* 2, e92.
- Hathaway RS, Nagda BRA, Gregerman SR (2002). The relationship of undergraduate research participation to graduate and professional education pursuit: an empirical study. *J Coll Stud Dev* 43, 614–631.
- Hunter A, Laursen S, Seymour E (2007). Becoming a scientist: the role of undergraduate research in students' cognitive, personal, and professional development. *Sci Educ* 91, 36–74.
- Hunter A-B, Weston T J, Laursen SL, Thiry H (2009). URSSA: evaluating student gains from undergraduate research in science education. *Counc Undergrad Res Q* 29, 15–19.
- Kardash CM (2000). Evaluation of an undergraduate research experience: perceptions of undergraduate interns and their faculty mentors. *J Educ Psychol* 92, 191–201.
- Kremer JF, Bringle RG (1990). The effects of an intensive research experience on the careers of talented undergraduates. *J Res Dev Educ* 24, 1–5.
- Kuh GD (2001). Assessing what really matters to student learning: inside the National Survey of Student Engagement. *Change* 33, 10–17.
- Laursen SL, Hunter A-B, Seymour E, Thiry H, Melton G (2010). *Undergraduate Research in the Sciences: Engaging Students in Real Science*, San Francisco: Jossey-Bass.
- Lave J, Wenger E (1991). *Situated Learning: Legitimate Peripheral Participation*, New York: Cambridge University Press.
- Lederman NG, Abd El Khalick F, Bell RL, Schwartz RS (2002). Views of nature of science questionnaire: toward valid and meaningful assessment of learners' conceptions of nature of science. *J Res Sci Teach* 39, 497–521.
- Leung W *et al.* (2010). Evolution of a distinct genomic domain in *Drosophila*: comparative analysis of the dot chromosome in *Drosophila melanogaster* and *Drosophila virilis*. *Genetics* 185, 1519–1534.
- Lopatto D (2004). Survey of undergraduate research experiences (SURE): first findings. *Cell Biol Educ* 3, 270–277.
- Lopatto D (2007). Undergraduate research experiences support science career decisions and active learning. *CBE Life Sci Educ* 6, 297–306.
- Lopatto D *et al.* (2008). Undergraduate research: Genomics Education Partnership. *Science* 322, 684.
- Lopatto D, Tobias S (2010). *Science in Solution: The Impact of Undergraduate Research on Student Learning*, Washington, DC: Council on Undergraduate Research.
- Lyman F (1996). The responsive classroom discussion: the inclusion of all students. In: *Mainstreaming Digest*, ed. AS Anderson, College Park: University of Maryland.

- Maton KI, Hrabowski FA III (2004). Increasing the number of African American PhDs in the sciences and engineering: a strengths-based approach. *Am Psychol* 59, 547–556.
- National Research Council (1996). *National Science Education Standards*, Washington, DC: National Academies Press.
- Netemeyer RG, Bearden WO, Sharma S (2003). *Scaling Procedures: Issues and Applications*, Thousand Oaks, CA: Sage.
- Olson S, Loucks-Horsley S (eds.) (2000). *Inquiry and the National Science Education Standards: A Guide for Teaching and Learning*, Washington, DC: National Academies Press.
- Packard BWL (2004). Mentoring and retention in college science: reflections on the sophomore year. *J Coll Stud Ret* 6, 289–300.
- Patel M, Trumbull D, Fox E, Crawford B (2009). Characterizing the inquiry experience in a summer undergraduate research program in biotechnology and genomics. Paper presented at the 82nd National Association of Research in Science Teaching (NARST) conference, held 17–21 April, Garden Grove, CA.
- Perna LW, Lundy-Wagner V, Drezner N, Gasman M, Yoon S, Bose E, Gary S (2009). The contribution of HBCUs to the preparation of African American women for STEM careers: a case study. *Res High Educ* 50, 1–23.
- Pope WH *et al.* (2011). Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. *PLoS One* 6, e16329.
- Porter SR, Rumann C, Pontius J (2011). The validity of student engagement survey questions: can we accurately measure academic challenge? *New Direct Institut Res* 150, 87–98.
- Quinn H, Schweingruber H, Keller T (eds.) (2011). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*, Washington, DC: National Academies Press.
- Rahm J, Miller H, Hartley L, Morre J (2003). The value of an emergent notion of authenticity: examples from two student/teacher-scientists partnership programs. *J Res Sci Teach* 40, 737–756.
- Rauchhorst WH, Czaja JA, Baxter Magolda M (2001). Measuring the impact of the undergraduate research experience on student intellectual development. Paper presented at Project Kaleidoscope Summer Institute, held 18–21 July, in Snowbird, UT.
- Rowland SL, Lawrie GA, Behrendorff JB, Gillam EM (2012). Is the undergraduate research experience (URE) always best? The power of choice in a bifurcated practical stream for a large introductory biochemistry class. *Biochem Mol Biol Educ* 40, 46–62.
- Russell CB, Weaver GC (2011). A comparative study of traditional, inquiry-based, and research-based laboratory curricula: impacts on understanding of the nature of science. *Chem Educ Res Pract* 12, 57–67.
- Russell SH, Hancock MP, McCullough J (2007). Benefits of undergraduate research experiences. *Science* 316, 548–549.
- Ryder J, Leach J, Driver R (1999). Undergraduate science students' images of science. *J Res Sci Teach* 36, 201–219.
- Savan B, Sider D (2003). Contrasting approaches to community based research and a case study of community sustainability in Toronto, Canada. *Local Environ* 8, 303–316.
- Seymour E, Hunter A-B, Laursen SL, DeAntoni T (2004). Establishing the benefits of research experiences for undergraduates in the sciences: first findings from a three year study. *Sci Educ* 88, 493–534.
- Shaffer CD *et al.* (2010). The Genomics Education Partnership: successful integration of research into laboratory classes at a diverse group of undergraduate institutions. *CBE Life Sci Educ* 9, 55–69.
- Singer SR, Hilton ML, Schweingruber HA (eds.) (2006). *America's Lab Report: Investigations in High School Science*, Washington, DC: National Academies Press.
- Siritunga D, Montero-Rojas M, Carrero K, Toro G, Vélez A, Carrero-Martínez FA (2011). Culturally relevant inquiry-based laboratory module implementations in upper-division genetics and cell biology teaching laboratories. *CBE Life Sci Educ* 10, 287–297.
- Smith MK, Wood WB, Adams WK, Wieman C, Knight JK, Guild N, Su TT (2009). Why peer discussion improves student performance on in-class concept questions. *Science* 323, 122–124.
- Stanton-Salazar RD (2011). A social capital framework for the study of institutional agents and their role in the empowerment of low-status students and youth. *Youth Soc* 43, 1066–1109.
- Szteinberg GA, Weaver GC (2013). Participants' reflections two and three years after an introductory chemistry course-embedded research experience. *Chem Educ Res Pract* 14, 23–35.
- 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.
- Thiry H, Laursen SL (2011). The role of student-advisor interactions in apprenticing undergraduate researchers into a scientific community of practice. *J Sci Educ Technol* 20, 771–778.
- Thiry H, Laursen SL, Hunter A-B (2010). What experiences help students become scientists? A comparative study of research and other sources of personal and professional gains for STEM undergraduates. *J High Educ* 82, 357–388.
- Thiry H, Weston TJ, Laursen SL, Hunter A-B (2012). The benefits of multi-year research experiences: differences in novice and experienced students' reported gains from undergraduate research. *CBE Life Sci Educ* 11, 260–272.
- Trochim WMK (2006). *Research Methods Knowledge Base: Measurement Validity Types*. www.socialresearchmethods.net/kb/measval.php (accessed 2 January 2014).
- Urban JB, Trochim W (2009). The role of evaluation in research—practice integration working toward the “golden spike.” *Am J Evaluat* 30, 538–553.
- Weaver GC, Russell CB, Wink DJ (2008). Inquiry-based and research-based laboratory pedagogies in undergraduate science. *Nat Chem Biol* 4, 577–580.
- Wei CA, Woodin T (2011). Undergraduate research experiences in biology: alternatives to the apprenticeship model. *CBE Life Sci Educ* 10, 123–131.
- Weiss C, Sosulski K (2003). *Quantitative Methods in Social Sciences e-Lessons: Measurement Validity and Reliability*. http://ccnmtl.columbia.edu/projects/qmss/measurement/validity_and_reliability.html (accessed 2 January 2013).
- Wenger E (2008). *Communities of Practice: A Brief Introduction*. <https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/11736/A%20brief%20introduction%20to%20CoP.pdf?sequence=1> (accessed 13 December 2013).
- WK Kellogg Foundation (2006). *Logic Model Development Guide*. www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide (accessed 13 December 2013).
- Wood WB (2003). Inquiry-based undergraduate teaching in the life sciences at large research universities: a perspective on the Boyer Commission Report. *Cell Biol Educ* 2, 112–116.
- Zeldin AL, Pajares F (2000). Against the odds: self-efficacy beliefs of women in mathematical, scientific, and technological careers. *Am Educ Res J* 37, 215–246.