

Essay

Modeling Course-Based Undergraduate Research Experiences: An Agenda for Future Research and Evaluation

Lisa A. Corwin,* Mark J. Graham,^{†‡} and Erin L. Dolan*

*Texas Institute for Discovery Education in Science, College of Natural Sciences, University of Texas at Austin, Austin, TX 78712; [†]Center for Teaching and Learning, Yale University, New Haven, CT 06520; [‡]Department of Psychiatry, School of Medicine, Yale University, New Haven, CT 06520

Submitted October 11, 2014; Revised November 25, 2014; Accepted December 10, 2014
Monitoring Editor: Mary Lee Ledbetter

Course-based undergraduate research experiences (CUREs) are being championed as scalable ways of involving undergraduates in science research. Studies of CUREs have shown that participating students achieve many of the same outcomes as students who complete research internships. However, CUREs vary widely in their design and implementation, and aspects of CUREs that are necessary and sufficient to achieve desired student outcomes have not been elucidated. To guide future research aimed at understanding the causal mechanisms underlying CURE efficacy, we used a systems approach to generate pathway models representing hypotheses of how CURE outcomes are achieved. We started by reviewing studies of CUREs and research internships to generate a comprehensive set of outcomes of research experiences, determining the level of evidence supporting each outcome. We then used this body of research and drew from learning theory to hypothesize connections between what students do during CUREs and the outcomes that have the best empirical support. We offer these models as hypotheses for the CURE community to test, revise, elaborate, or refute. We also cite instruments that are ready to use in CURE assessment and note gaps for which instruments need to be developed.

INTRODUCTION

Course-based undergraduate research experiences (CUREs) are gaining attention as an effective way to engage students in doing research early in their college careers. CUREs offer several advantages over traditional lab courses and research internships. They involve many undergraduates in science research at one time, and all students who enroll in a course are able to participate (Corwin Auchincloss *et al.*, 2014). Studies of students who complete CUREs report many of the gains associated with undergraduate participation in

research internships (Shaffer *et al.*, 2010; Rowland *et al.*, 2012; Jordan *et al.*, 2014). For example, similar to studies of undergraduate research experiences, studies of CUREs report student gains in research skills, self-efficacy, and intent to persist in science (Lopatto *et al.*, 2008; Shaffer *et al.*, 2010; Harrison *et al.*, 2011; Rowland *et al.*, 2012; Corwin Auchincloss *et al.*, 2014; Jordan *et al.*, 2014).

Most research and evaluation of CUREs have focused on documenting student outcomes. What remains largely unaddressed is which aspects of CUREs lead to desirable student outcomes. In other words, what are the causal mechanisms underlying the efficacy of CUREs? We hypothesize that CUREs allow students to participate in a unique combination of activities that result in progressive achievement of diverse cognitive, psychosocial, and behavioral outcomes. We posit that using experimental concepts to model pathways, or directional relationships between student activities and outcomes, will allow us to identify the critical features that should be included in the design of all CUREs and gain insight into the latitude we have in effectively implementing CUREs.

Logic models are often used to illustrate relationships between program activities and resulting outcomes and define

CBE Life Sci Educ March 2, 2015 14:es1

DOI:10.1187/cbe.14-10-0167

Address correspondence to: Lisa A. Corwin (lisa.c.a@utexas.edu).

© 2015 L. A. Corwin *et al.* CBE—Life Sciences Education © 2015 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.

questions for evaluation and research (Westat *et al.*, 2010). However, the simplified lists used in logic models do not capture relationships and feedback loops or the high level of impact certain activities may have. The systems approach to evaluation (Urban and Trochim, 2009; Urban *et al.*, 2014) expands on a traditional logic model by hypothesizing direct and directional relationships among activities and outcomes that evolve over time. Specifically, the systems approach to evaluation does the following:

- Allows for description and visualization of the many diverse connections among activities and outcomes
- Displays multiple pathways students may take to achieve a single outcome
- Accommodates the complexity of education programs by exposing feedback loops and multiple connections among outcomes
- Reveals which outcomes are likely to be most informative by pinpointing where pathways converge

This systems approach to evaluation has been used to design evaluation plans for several existing or planned programs (Urban and Trochim, 2009). We used this approach to generate a broad model of what is known about CURE instruction based on research to date and to identify leverage points at which more information or measures about CUREs are needed. We propose that the resulting model can be used or adapted to generate evaluation plans and an agenda for research on CUREs.

As context for identifying causal mechanisms of CUREs—what makes CUREs work for students—it is useful to consider theoretical perspectives on learning. Social learning theory (Bandura, 1971; Vygotsky, 1978), which has both cognitive and social elements, is particularly useful for considering what makes CUREs beneficial for students. This theory frames learning as a cognitive process, reliant on mental processing and construction, which occurs in a social context through observing and interacting with others. Situated-learning theory is a form of social learning theory that emphasizes the importance of situating learning in an authentic activity, context, and culture (Brown *et al.*, 1989; Lave and Wenger, 1991). For the most part, CUREs are designed to align with the tenets of situated learning. Students do the work that scientists do (e.g., ask questions, design studies, collect and analyze data, build models) in the context of a real scientific problem or question, in which the solution or answer is unknown.

A hallmark of situated learning is “legitimate peripheral participation,” meaning the learner does tasks that experts consider meaningful (“legitimate”) to the work of the discipline. The tasks become increasingly important, moving along a continuum from peripheral to central, as the learner develops expertise (Lave and Wenger, 1991). Students within CUREs are legitimate participants in scientific research, because their actions contribute to achievement of research goals. However, students generally do not perform more central tasks that determine the overall direction and scope of research. For example, in many CUREs, instructors do the central work of posing overarching research questions, which helps steer students in scientifically fruitful directions. Students then do the very real (legitimate) but more peripheral work of collecting and analyzing data to answer those questions. As students develop more expertise, they

take on increasingly central roles, for example, by designing or choosing methods and eventually posing research questions themselves.

Working with more expert individuals is another important element of situated learning, because experts model how to do the work, provide feedback about how to improve, and validate the legitimacy of students’ experiences and accomplishments. Situated-learning theory emphasizes the importance of immersion in real culture—in the case of CUREs, the culture of science, which is characterized by scientific ways of thinking, behaving, and working. Through immersion in the culture of science, students not only have opportunities to see science and scientific thinking in action but also to develop in terms of their scientific identity and sense of belonging to the broader scientific community. CUREs are thought to more closely reflect the culture of science through immersion in the process of science when compared with traditional lab-learning experiences (Spell *et al.*, 2014). According to situated-learning theory, we would predict that participating in CUREs would lead to students developing a stronger scientific identity and sense of belonging to the scientific community. We would also expect to observe positive relationships between the degree to which students interact with more expert individuals and the positive outcomes they achieve.

To date, most program evaluation of CUREs has focused on high-stakes outcomes, such as student completion of a science, technology, engineering, and mathematics (STEM) major or matriculation into a science graduate program. Yet the causal elements of these outcomes have not been systematically identified (Corwin Auchincloss *et al.*, 2014). Although it is important to continue to measure these outcomes, this should be paired with efforts to understand how they are achieved. Building and testing models, such as the ones we present here, will help to direct and improve research and evaluation of CUREs as the biology education community shifts focus beyond “What works?” to “How does it work?” Thus, the purpose of this work is to provide the biology education community with a working model that can be used, adapted, and revised to drive future research on CUREs and assist in CURE evaluation efforts. Our model is intended to show how CUREs “work”—in other words, what it is that students do during CUREs that may lead to the outcomes they experience. To build this model, we:

1. reviewed relevant studies to determine what is known about student outcomes from undergraduate research experiences and how they have been measured;
2. evaluated the studies related to each outcome to determine the extent of empirical support for each outcome—in other words, whether the outcome is probable, possible, or proposed;
3. connected learning activities with the outcomes they are likely to influence to produce several small models of how CUREs work (“mini-models”);
4. integrated our mini-models into a single, large model to show student progression from activities to short-, medium-, and long-term outcomes; and
5. identified points at which multiple pathways converge or diverge, called “hubs” (see Urban and Trochim, 2009), which we identify as key points for future research and evaluation of CUREs.

METHODS

Step 1. Review of Relevant Literature to Identify Potential CURE Outcomes

We started by assembling a comprehensive set of papers describing studies of CUREs and undergraduate research internships to identify the range of student outcomes. We included studies of internships, because a number of CUREs were developed as a mechanism for involving students in research when sufficient internships were not available (Dolan *et al.*, 2008), and CURE outcomes have been compared with internship outcomes (Lopatto *et al.*, 2008; Shaffer *et al.*, 2010). We included literature cited in a recent report on CURE assessment (Corwin Auchincloss *et al.*, 2014) and in two comprehensive reviews of undergraduate research (Seymour *et al.*, 2004; Laursen *et al.*, 2010). We also searched Google Scholar using the terms “course based undergraduate research experience,” “undergraduate research experience,” “undergraduate research internship,” and “class undergraduate research” for additional studies published within the past 5 yr. All studies included in our analysis met three criteria. They 1) were published within the past 25 yr, 2) presented data gathered in undergraduate settings, and 3) actively examined student outcomes of a CURE or a research internship (i.e., they had elements of experimental design and provided more than anecdotal or descriptive evidence of outcomes).

For the purposes of this paper, we defined a CURE as “a course in which students are expected to engage in science research with the aim of producing results that are of interest to a scientific community.” This definition is purposely broad, since the processes of science research conducted in laboratory courses vary widely, and we have not yet reached consensus on the essential elements that constitute a CURE or make CUREs successful. Laboratory learning experiences described in these studies were deemed CUREs if they reported that the students performed externally relevant scientific research at any point during the course and for any duration. We identified 14 studies on CUREs that met these criteria and 25 studies on undergraduate research internships. In general, the duration of research performed in the identified CURE studies spanned the majority of the course, involved students in multiple science practices (e.g., collecting data, analyzing data, reporting on data), and involved a minimum of 15 students.

We then generated a list of potential CURE outcomes based on evidence or hypotheses presented in this literature sample. We tested the comprehensiveness of this literature sample by searching for additional studies using the same terms described above. We searched two digital libraries (ERIC and JSTOR) and a crowd-sourced database (Mendely) and reviewed the resulting studies for additional outcomes. We found one new study on CUREs, two studies on internships, and no new outcomes. Thus, we deemed our literature sample sufficient for building models of CUREs. When conducting step 2, we included all studies that fit the four criteria noted earlier.

Step 2. Analysis to Determine the Level of Empirical Support

To determine which outcomes had sufficient empirical support to be included in a model, we designated outcomes

as “probable,” “possible,” and “proposed,” using a framework similar to that used by Grayson *et al.* (2001). Probable outcomes were 1) investigated in a minimum of three studies, 2) measured in at least three different student populations (i.e., groups of students), 3) measured in at least three different courses or curricula, and 4) assessed using at least two different methods or instruments. Possible outcomes were 1) investigated in a minimum of two studies, 2) investigated in two different populations, 3) measured in at least one course or curriculum, and 4) assessed using at least one method. Proposed outcomes were investigated only in a single instance or were supported by learning theory but were not present in the literature sample. Purely descriptive accounts and anecdotes were not used to support these designations. We chose to represent outcomes using discrete ordinal designations in order to determine which outcomes had sufficient support to include in our models. We did not identify any outcomes as certain in our analysis, because current research on CUREs is not sufficient for this designation. We used probable or possible outcomes in subsequent modeling steps.

Step 3. Alignment of Activities and Outcomes to Generate Mini-Models

We connected activities to outcomes to generate several mini-models (Figures 1–3). Following on the systems approach to evaluation (Urban and Trochim, 2009), we parsed outcomes into short, medium, and long term. We defined short-term outcomes as those that can be achieved immediately during a CURE and medium-term outcomes as those that result primarily from achievement of short-term outcomes and occur later, at the end of a CURE or after CURE participation. We defined long-term outcomes as those that result from short- and medium-term outcomes and can only be measured after the CURE is completed. These are often highly valued goals of science education and research experiences in particular, such as development of students’ scientific identity, student persistence in science, and increased public science literacy. All modeling was performed using the Netway tool for program evaluation and planning (Cornell Office of Research and Evaluation [CORE], 2009)

To support the presence and direction of connections in our models, we used social learning theory (described above), investigations of STEM retention outcome relationships (e.g., Estrada *et al.*, 2011), and hypotheses proposed in the CURE literature. We connected activities to short-term outcomes in mini-models and then connected short-to medium-term outcomes. Collectively, the mini-models include the majority of the probable or possible outcomes of CUREs (Table 1 and Figures 1–3).

Step 4. Integration of Mini-Models into a Comprehensive Model

Using the Netway (CORE, 2009), we combined the mini-models into a single large model that depicts the relationships among all activities and outcomes and expands the model to include long-term outcomes (Figure 4). To accomplish this, we used learning theory to hypothesize additional connections among short- and medium-term outcomes and between medium- and long-term outcomes (Brown *et al.*, 1989; Lave and Wenger, 1991).

Table 1. Support for CURE outcomes based on a review of relevant CURE literature^a

	Outcome	CURE References
Probable	Increased content knowledge	Lopatto <i>et al.</i> , 2008; Shaffer <i>et al.</i> , 2010, 2014; Siritunga <i>et al.</i> , 2011; Brownell <i>et al.</i> , 2012; Rowland <i>et al.</i> , 2012; Jordan <i>et al.</i> , 2014; Kloser <i>et al.</i> , 2013
	Increased analytical skills	Shaffer <i>et al.</i> , 2010, 2014; Siritunga <i>et al.</i> , 2011; Bascom-Slack <i>et al.</i> , 2012; Brownell <i>et al.</i> , 2012; Hanauer <i>et al.</i> , 2012; Alkahaer and Dolan, 2014; Jordan <i>et al.</i> , 2014
	Increased self-efficacy	Drew and Triplett, 2008; Lopatto <i>et al.</i> , 2008; Shaffer <i>et al.</i> , 2010, 2014; Siritunga <i>et al.</i> , 2011; Kloser <i>et al.</i> , 2013; Jordan <i>et al.</i> , 2014
	External validation from a science community	Hatfull <i>et al.</i> , 2006; Lopatto <i>et al.</i> , 2008; Caruso <i>et al.</i> , 2009; Shaffer <i>et al.</i> , 2010, 2014; Jordan <i>et al.</i> , 2014
	Persistence in science	Drew and Triplett, 2008; Harrison <i>et al.</i> , 2011; Hanauer <i>et al.</i> , 2012; Bascom-Slack <i>et al.</i> , 2012; Brownell <i>et al.</i> , 2012; Jordan <i>et al.</i> , 2014; Shaffer <i>et al.</i> , 2014
	Increased technical skills	Drew and Triplett, 2008; Shaffer <i>et al.</i> , 2010; Jordan <i>et al.</i> , 2014; Rowland <i>et al.</i> , 2012
	Career clarification	Drew and Triplett, 2008; Harrison <i>et al.</i> , 2011; Shaffer <i>et al.</i> , 2014
Possible	Increased project ownership	Shaffer <i>et al.</i> , 2010; Hanauer <i>et al.</i> , 2012; Alkahaer and Dolan, 2014
	Increased communication skills	Lopatto <i>et al.</i> , 2008; Jordan <i>et al.</i> , 2014; Shaffer <i>et al.</i> , 2014
	Increased motivation in science	Shaffer <i>et al.</i> , 2010, 2014; Alkahaer and Dolan, 2014
	Increased collaboration skills	Shaffer <i>et al.</i> , 2010, 2014
	Increased tolerance for obstacles	Jordan <i>et al.</i> , 2014; Shaffer <i>et al.</i> , 2014
	Increased sense of belonging to a larger community	Jordan <i>et al.</i> , 2014; Shaffer <i>et al.</i> , 2014
	Enhanced science identity	Hanauer <i>et al.</i> , 2012; Alkahaer and Dolan, 2014
Increased positive interaction with peers	Shaffer <i>et al.</i> , 2010; Alkahaer and Dolan, 2014	
Proposed	Increased access to faculty interaction	Alkahaer and Dolan, 2014
	Increased access to mentoring functions	Hanauer <i>et al.</i> , 2012
	Enhanced understanding of the nature of science	Russell and Weaver, 2011
	Development of self-authorship	Alkahaer and Dolan, 2014

^aGreen shading indicates probable outcomes, yellow shading indicates possible outcomes, and gray shading indicates proposed outcomes.

Step 5. Identification of “Hubs”

Hubs are points, or “nodes,” in a model at which multiple paths converge (Urban and Trochim, 2009). Hubs are important, because many pathways must pass through a hub for longer-term outcomes to be achieved. Thus, students would be expected to achieve at least one hub outcome in order to achieve longer-term outcomes. For the present study, we designated as hubs those outcomes that had a minimum of six relationships, or directional connections, with other outcomes or activities (Figure 4).

RESULTS

Our review and analysis revealed eight probable, seven possible, and four proposed outcomes (Table 1). More than half of the outcomes that may result from CURE participation are only possible or proposed in the CURE literature.

Level of Empirical Support for Outcomes

Probable Outcomes. Increases in content knowledge, technical skills, analytical skills, scientific self-efficacy, project ownership, and career clarification are all probable outcomes, according to our criteria. In addition, two studies tracked students after their participation in CUREs and

found that they pursued science graduate degrees or science-related careers at a higher rate than other science majors (Bascom-Slack *et al.*, 2012; Hanauer *et al.*, 2012). A third study showed that CURE students persisted in science majors at a higher rate than students who completed other lab courses (Jordan *et al.*, 2014). In three additional studies, students reported increased intentions to continue doing research, including pursuing graduate education. This is important, because several researchers have found that students’ educational aspirations are one of the strongest predictors of enrollment in a graduate degree program (Sewell *et al.*, 1969; Heller, 2001; Mullen *et al.*, 2003; Walpole, 2003; Nevill and Chen, 2007; Eagan *et al.*, 2013). We grouped the outcomes of continuing in a science major, entering a graduate program, or pursuing a science career, including students’ intentions to do so and their actual behavior, under the broad heading of “persistence in science.” A number of studies showed evidence of validation of students and their contributions by a broader scientific community, which we call “external validation.” Authorship or acknowledgment in a peer-reviewed journal, presentation at a professional conference, and acceptance of results into a national database all constitute forms of external validation reported in CURE studies.

These results raise several points for consideration. First, probable outcomes with the best support were described

in seven unique instances (i.e., different courses or distinct groups of students). Although this met our criteria, future work should focus on investigating the prevalence of these outcomes across a broader set of CUREs that represent the diversity of instruction and students. Second, in a few CURE studies, skill gains were measured through tests, observation, or instructor evaluations of student work. In most, however, students reported gains in their skills. This can be problematic, because novices tend to overestimate their own knowledge and skills, a phenomenon described as the Dunning-Kruger effect (Boud and Falchikov, 1989; Falchikov and Boud, 1989). More direct measurements of technical and analytical skill development in CUREs are needed to corroborate current findings. Third, certain CUREs may be subject to the same recruitment or selection bias as research internships, in that they only involve a small group of high-achieving or research-interested students (Brownell *et al.*, 2013). Moving forward, it will be important to study the impact of CURE instruction using approaches that control for student-level variables and nonrandom assignment (Theobald and Freeman, 2014; Beck and Bliwise, 2014), such as propensity score matching (e.g., Schultz *et al.*, 2011) or regression discontinuity design (e.g., DesJardins *et al.*, 2010, 2014).

Possible Outcomes. We categorized increased communication skills, collaboration skills, motivation to pursue science learning, and enhanced science identity as possible outcomes of CUREs, because they were documented in at least two unique instances (i.e., different courses and different student groups). We classified descriptions of successful peer-peer instruction (Shaffer *et al.*, 2010) and productive scientific discourse with peers (Alkahrer and Dolan, 2014) as “positive interaction with peers” and also categorized this as possible. Two studies reported increases in the following two outcomes: 1) tolerance for obstacles and 2) an increased sense of belonging to a larger community (Shaffer *et al.*, 2010; Jordan *et al.*, 2014). Because these studies examined two distinct CUREs, we categorized these outcomes as possible. However, both outcomes were measured with a single question each, using a single instrument for which validity and reliability information is not available. Moving forward, it will be important to establish the validity and reliability of outcome measures and demonstrate outcomes with multiple kinds of evidence and within many different kinds of CUREs.

Proposed Outcomes. In our analysis, we found only single instances of increased access to faculty (Alkahrer and Dolan, 2014), increased access to mentoring functions (Hanauer *et al.*, 2012), enhanced understanding of the nature of science (Russell and Weaver, 2011), and development of self-authorship (Alkahrer and Dolan, 2014). These are notable shortcomings, because the development of meaningful relationships with faculty, access to mentoring, and self-authorship are predictors of persistence in science and career decision making, especially for students who are underrepresented in science (Maton *et al.*, 2000; Packard, 2004; Creamer and Laughlin, 2005; Eagan *et al.*, 2010, 2011). More evidence is needed to elucidate whether these outcomes are replicable in particular CUREs or in CUREs in general.

Mini-Models: Connecting Activities to Outcomes

We used both probable and possible outcomes to construct mini-models of connections between activities and outcomes of CUREs. Although there is theoretical support for many of the relationships, the connections we show indicate that most have not been empirically investigated. We intend this set of models to illustrate how to go about pathway modeling of CURE instruction; they are not the only models possible for CUREs or as a comprehensive set of such models.

We chose to focus on six activities that students typically engage in during CURE instruction: reading and evaluating science literature, selecting or designing methods, collecting novel data, analyzing results, working collaboratively, and presenting results outside class (Hatfull *et al.*, 2006; Drew and Triplett, 2008; Caruso *et al.*, 2009; Shaffer *et al.*, 2010; Siritunga *et al.*, 2011; Brownell *et al.*, 2012; Hanauer *et al.*, 2012; Jordan *et al.*, 2014). We do not expect that all CUREs will include all of these activities or be limited to them. If this model is adapted for use in research and evaluation of CUREs, it should be tailored accordingly.

Knowledge and Skills Mini-Model. This model depicts how students develop content knowledge and hone their technical and analytical skills when they read and evaluate scientific literature, collect data, and analyze results (Figure 1; Shaffer *et al.*, 2010; Brownell *et al.*, 2012; Rowland *et al.*, 2012; Siritunga *et al.*, 2011; Kloser *et al.*, 2013). Increases in knowledge and skills lead to improved scientific self-efficacy (Thiry and Laursen, 2011; Hanauer *et al.*, 2012) and, ultimately, increased motivation to learn more science, which further improves science efficacy (Graham *et al.*, 2013).

Communication and Collaboration Mini-Model. This model depicts how, when students work collaboratively

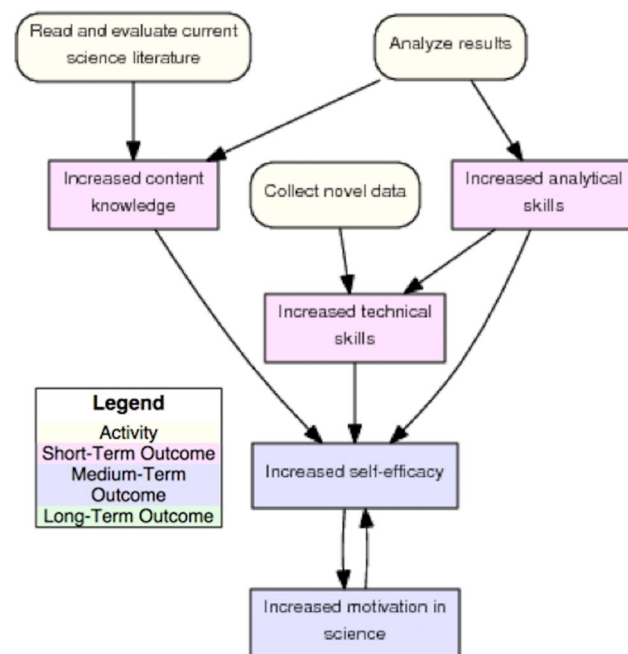


Figure 1. Knowledge and skills mini-model. Arrows represent positive directional relationships between activities and outcomes.

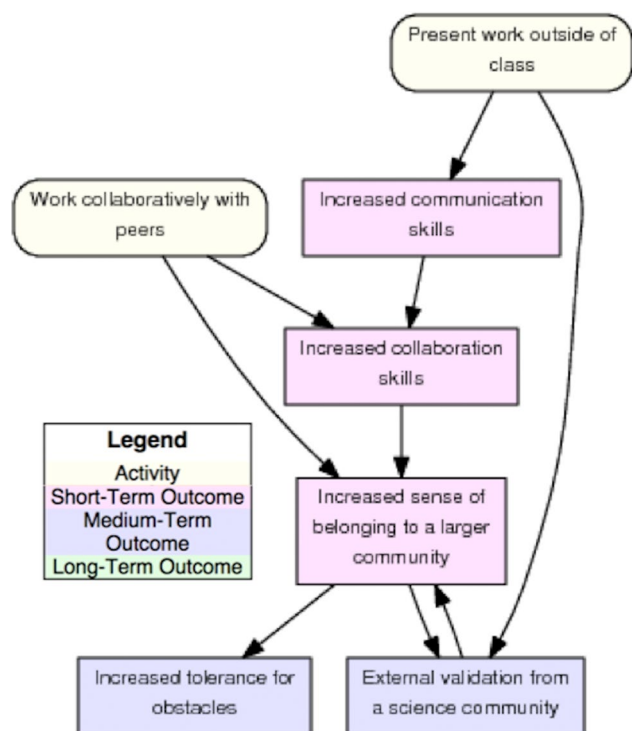


Figure 2. Communication and collaboration mini-model. Arrows represent positive directional relationships between activities and outcomes.

and communicate about their work outside class (Figure 2; Caruso *et al.*, 2009; Bascom-Slack *et al.*, 2012; Shaffer *et al.*, 2014; Jordan *et al.*, 2014), they improve their communication and collaboration skills (Seymour *et al.*, 2004; Laursen *et al.*, 2010; Jordan *et al.*, 2014; Shaffer *et al.*, 2014). The more students collaborate and communicate, the more they feel they belong to a larger community (Alexander *et al.*, 1998), such as a lab community (Barlow and Villarejo, 2004) or a science learning community (Siritunga *et al.*, 2011; Jordan *et al.*, 2014). Theoretically, students who have opportunities to present to members of a scientific community or who develop professional and personal networks that connect them with a broader community would have increased opportunities for external validation (e.g., publishing; Lave and Wenger, 1991; Wenger, 1998). This external validation further solidifies a student’s role as a community member (Lave and Wenger, 1991). Through feedback, support, and role modeling from more experienced scientists, students can learn to tolerate obstacles and failure and develop the temperament necessary for research work (Thiry and Laursen, 2011; Seymour *et al.*, 2004).

Ownership Mini-Model. This model depicts how students develop a sense of project ownership when they have agency to design their own studies, choose experimental methods, and collect data of interest to them or their community (Figure 3; Hanauer *et al.*, 2012). A growing sense of ownership increases students’ tolerance for obstacles and perseverance, motivating them to complete their projects even in the face of challenges (Ward *et al.*, 2002; Laursen *et al.*, 2010; Hanauer *et al.*, 2012; Alkahrer and Dolan, 2014). This has also

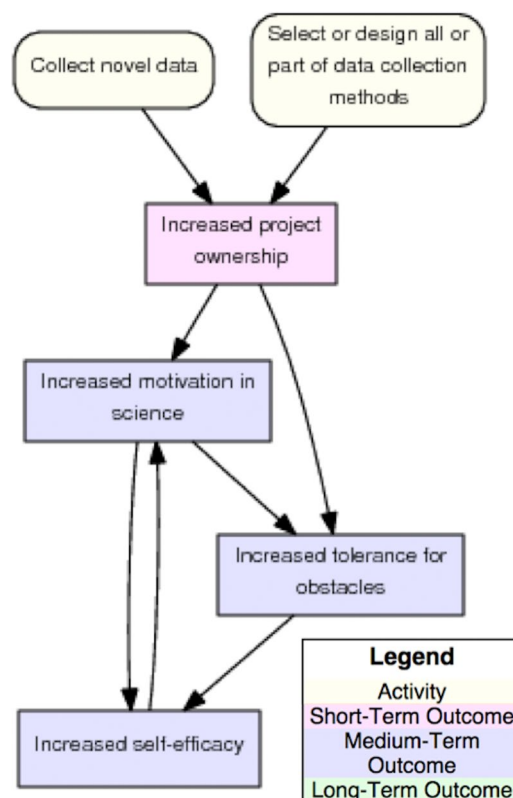


Figure 3. Ownership mini-model. Arrows represent positive directional relationships between activities and outcomes.

been conceptualized as “grit” (Duckworth *et al.*, 2007). When students experience success by overcoming obstacles in their research, they develop a greater sense of their scientific self-efficacy, which increases motivation. These outcomes could operate as a feedback loop. When students experience success in overcoming obstacles, their self-efficacy improves, which increases their motivation, which increases likelihood of success in overcoming new obstacles, and so on (Ward *et al.*, 2002, Thiry *et al.*, 2012, Graham *et al.*, 2013).

An Integrated Large Model: Progress toward Long-Term Outcomes

We combined the three mini-models to construct a large CURE model representing our current understanding of how students could achieve outcomes by participating in CUREs (Figure 4). The mini-models are interrelated, because single activities can lead to multiple outcomes and achieving short- and medium-term outcomes is likely to be important for realizing long-term outcomes. To create the large CURE model, we formed two additional connections between short-term outcomes: 1) students who work collaboratively with peers make improvements in their technical skills by receiving modeling and feedback on how to perform particular tasks (Lave and Wenger, 1991); and 2) students who feel ownership over their projects develop an increased sense of belonging to a science community, since they see the science projects as extensions of themselves (Wiley, 2009; Hanauer *et al.*, 2012).

We also formed five new connections between medium- and long-term outcomes and three new connections among

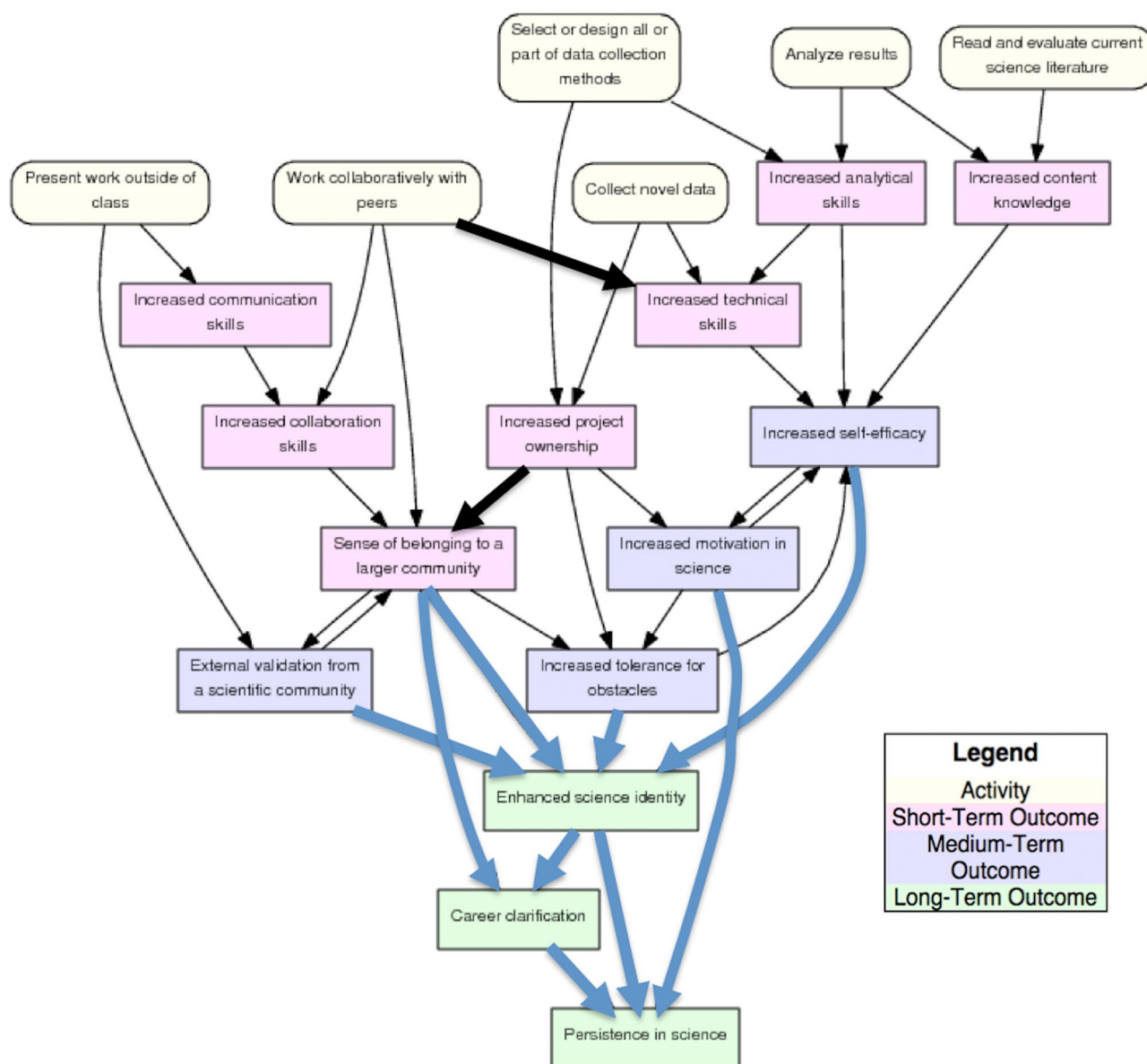


Figure 4. Large CURE model. Arrows represent positive directional relationships between activities and outcomes. Bold black arrows indicate new connections between activities and short-term outcomes in the mini-models. Bold blue arrows indicate new connections between medium- and long-term outcomes.

long-term outcomes to form the new pathways depicted in Figure 4. We designated enhanced science identity, career clarification, and persistence in science as long-term outcomes, since they are likely to occur late in CURE participation or after CURE participation. Here, we define “persistence” as students staying in a science track 1 yr after they participate in a CURE. Thus, persistence includes staying in a science major, matriculating into a graduate program, or pursuing a career in science depending on when the CURE falls in a student’s academic trajectory. Many medium-term outcomes have potential to enhance students’ science identity. As students grow in their sense of belonging to a larger scientific or laboratory community and as they receive validation from that community, they are likely to further identify as scientists and decide

whether a career in science is of interest (Lave and Wenger, 1991). Tolerance for obstacles is a broadly recognized characteristic of a scientific disposition; self-recognition of this characteristic also enhances science identity (Seymour *et al.*, 2004; Thiry and Laursen, 2011; Thiry *et al.*, 2012). Increased scientific self-efficacy increases motivation and scientific identity, both of which influence persistence in science (Harrison *et al.*, 2011; Estrada *et al.*, 2011; Adedokun *et al.*, 2013).

We expect other factors, such as lower social barriers and increased environmental support (see Lent *et al.*, 1994; Estrada *et al.*, 2011; Graham *et al.*, 2013), to mediate and contribute to identity and persistence in science. Despite support for these hypotheses, they are not modeled, because they have not yet been established as probable or possible CURE outcomes.

Identification of Hubs, Reciprocal Relationships, and Feedback Loops

A distinct advantage of the systems approach to evaluation is the process of identifying hubs, reciprocal relationships, and feedback loops. Our model shows that scientific identity is a hub, because multiple activities and short- and medium-term outcomes converge on scientific identity, and scientific identity is on multiple paths to the highly desired outcome of persistence in science (American Association for the Advancement of Science, 2011; President's Council of Advisors on Science and Technology, 2012). When evaluation resources are limited, evaluation of hubs should be prioritized. For example, if a CURE instructor or designer wants to know whether his or her CURE is effective, it may be best to prioritize measuring changes in students' scientific identity over changes in content knowledge. Similarly, when resources for instrument development are limited, developing valid and reliable ways of measuring hub outcomes should be prioritized.

In the large model, we identified what we believe to be reciprocal relationships and feedback loops, which we posit may be useful for setting evaluation priorities. We define a "reciprocal relationship" as one in which outcomes are linked via mutual positive relationships. For example, self-efficacy and motivation are likely to be reciprocally linked. As a student develops greater self-efficacy, he or she is likely to be more motivated, which in turn will further his or her sense of self-efficacy. There is good consensus on how to define self-efficacy, and there are published measures of scientific and research self-efficacy (Lent *et al.*, 1994; Chemers *et al.*, 2010; Estrada *et al.*, 2011). A similar consensus does not exist about what constitutes student motivation in science. Thus, priority should be placed on evaluating only one of two reciprocally related activities or outcomes, with a mind to which may be most easily or accurately measured.

We propose that feedback loops occur when an outcome has an indirect effect on itself via downstream outcomes. Feedback loops are distinct from reciprocal relationships, because feedback loops operate over time and should be empirically identifiable. For example, increased self-confidence should lead to increased motivation, which will help students overcome obstacles. This will further increase students' confidence (Graham *et al.*, 2013). Thus, if students have the opportunity to encounter and fix problems, we should continue to see incremental increases in self-efficacy with each new obstacle they overcome. This can be empirically tested in a longitudinal study. Longitudinal studies will be useful for evaluating longer pathways to achieving outcomes and for identifying long-term feedback loops. For example, designing a project should lead to an increased sense of ownership, which in turn leads to an enhanced sense of belonging to a larger community, enhanced scientific identity, and, ultimately, completion of a science major and possible pursuit of a graduate degree, which will afford more opportunity for identity development and so on. However, this may take years to assess. When time and resources are limited, priority should be placed on measuring short- and medium-term outcomes. If these are not achieved, then effort should be invested in improving CURE instruction to better achieve short- and

medium-term outcomes rather than the labor-intensive, long-term tracking of students.

DISCUSSION

The systems approach to evaluation emphasizes the importance of developing a model of a program before selecting which outcomes to measure. We can use models to identify outcomes of particular interest. Urban and Trochim (2009) proposed three categories of outcomes to prioritize: 1) "hubs," which are highly connected diagnostic outcomes; 2) "low-hanging fruit," which are easily measurable outcomes; and 3) "pinnacle" outcomes, which are important for stakeholder interests (e.g., funding agencies, institutional priorities, accreditation) or necessary for continuing the program. We should also consider the availability of existing measures (protocols, rubrics, surveys, etc.) in determining which outcomes to assess and when to assess them. We can designate early, middle, and late evaluation phases that are developed based on collective consideration of this information, the timeline for CURE development, and the timeline for when students are expected to realize particular outcomes.

We offer three phases for evaluation of CUREs to illustrate this process (Figure 5). We discuss which outcomes constitute low-hanging fruit, pinnacle outcomes, and hubs in each phase. We designed phases around when outcomes are likely to be achieved, taking into account the ease of assessment. We recommend that each CURE evaluation be designed in program-specific phases that consider particular CURE activities, stakeholder interests, programmatic goals, and evaluation resources.

Early-Phase Evaluation

When CURE instructors or designers are first implementing and evaluating their CUREs, this is the place to start. This phase focuses on whether activities are resulting in short-term outcomes. Measures exist for many of these outcomes (low-hanging fruit), including assessment of analytical research skills (e.g., developing hypotheses, designing experiments, analyzing data; Feldon *et al.*, 2011; Sirum and Humburg 2011; Brownell *et al.* 2014; Dasgupta *et al.*, 2014; Deane *et al.*, 2014), project ownership (Hanauer and Dolan, 2014), and the hub of self-efficacy (Chemers *et al.*, 2010; Estrada *et al.*, 2011). In addition, these outcomes should be measurable during or immediately after CURE instruction.

Middle-Phase Evaluation

This phase focuses on assessment of the communication and collaboration elements of CUREs and includes one hub: becoming part of a larger community. This phase will require the development and testing of new instruments, because, to the best of our knowledge, there are no measures of communication or collaboration skills or of students' sense of belonging to a science community that have been tested with CURE students, although there are tools that could be adapted for these purposes. For example, Sevian and Gonsalves (2008) offer a rubric for measuring the effectiveness of scientific explanations, but it was developed for use with graduate students explaining their work to middle-school students. Similarly, there are general measures of undergraduate students' sense

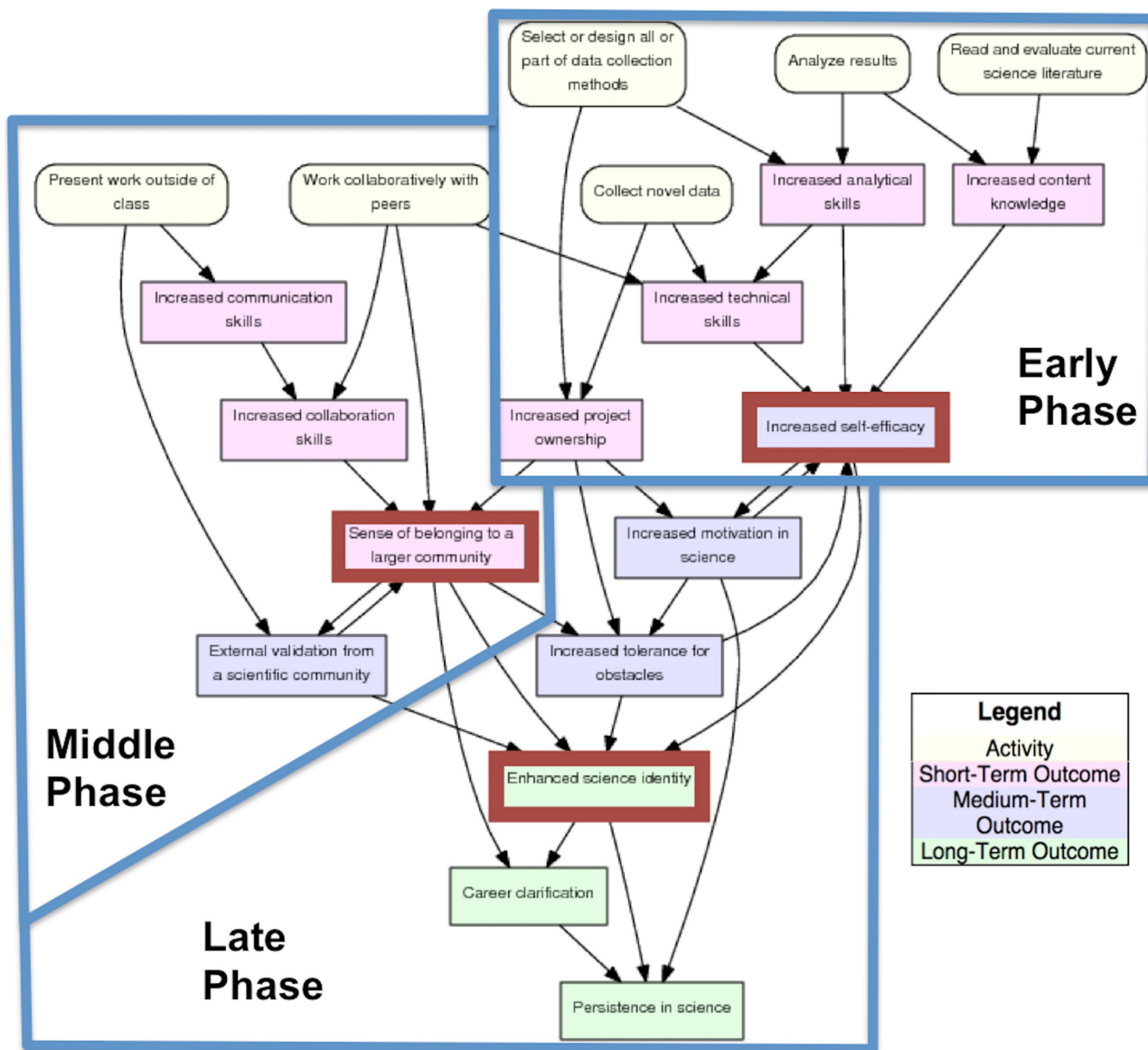


Figure 5. Large CURE model showing phases and hubs. Arrows represent positive directional relationships between activities and outcomes. Blue boxes indicate separate evaluation phases. Red borders surrounding outcomes indicate “hub” outcomes.

of belonging or sense of community and the relationship between these and students’ intended or actual persistence in college (Bollen and Hoyle, 1990; Chipuer and Pretty, 1999; Hausmann *et al.*, 2007, 2009). Again, these tools need to be adapted and tested to demonstrate their usefulness for CURE assessment. Some of the outcomes in the middle phase also take longer to achieve. For instance, it may take time to establish a mechanism through which students can present work outside class or it may take multiple iterations of a CURE to yield sufficient data for publication in a science journal.

Late-Phase Evaluation

Late-phase evaluation focuses on medium- and long-term outcomes of CUREs, including scientific identity (Chemers *et al.*, 2010; Estrada *et al.*, 2011), which is both a hub and a

low-hanging fruit. As with the middle-phase example, we could not find established, valid, and reliable measures of science student motivation, tolerance of obstacles, or career clarification. There are measures that have some potential to be adapted for CURE assessment. For example, the Science Motivation Questionnaire, developed for use with non science majors, might be adaptable for use with majors, depending on whether and how motivation differs between majors and nonmajors (Glynn *et al.*, 2009). The “grit scale” is a general measure of perseverance and passion for long-term goals that could be adapted to be more science specific (Duckworth *et al.*, 2007). The Undergraduate Research Student Self-Assessment includes a career clarification scale that is well grounded in qualitative research but for which validity and reliability information have not been published (Hunter *et al.*, 2009). Although the “pinnacle” outcome of

persistence in science is the long-term outcome featured here, there are other long-term outcomes likely to be of interest and significant value, such as scientific literacy and understanding of the nature of science.

There is a temptation to focus on assessing pinnacle outcomes because of their value to stakeholders and because they are often more compelling than more immediate outcomes. Yet it is important to keep in mind that pinnacle outcomes, such as persistence in science and understanding the nature of science, may take longer to achieve and may only be achieved for a subset of CUREs or by certain students or following multiple CURE experiences. It is also exceedingly difficult to design and conduct a study that shows a causal relationship between an instructional experience and student outcomes, and this difficulty is compounded when the outcomes take an extended time to realize. For example, we included persistence in science as an outcome with sufficient evidence to include in models used to study CUREs moving forward. However, as noted above, the studies we found did not control for the extent to which students may already be planning to persist in science (self-selection). In addition, these studies did not make use of established methods for designing and interpreting quasi-experimental studies in a way that provides definitive evidence that participation in CUREs improves persistence in science. The CURE community should consider a handful of larger-scale collaborations aimed at examining the relationship between CURE activities and desired long-term outcomes rather than expecting individual instructors to collect data that can be used to answer questions about long-term impacts of CURE instruction.

Each of these models, as well as each pair of nodes and the relationship between them, is a testable hypothesis about *how* CUREs work. For example, we can investigate whether students who develop an improved sense of self-efficacy are more tolerant of obstacles, whether students who develop technical skills also develop an increased sense of self-efficacy, and whether students who collect novel data improve their technical skills. We expect that other aspects of CUREs not addressed here, such as the duration of the experience, will influence these relationships. It is also important to investigate how student differences, such as gender, major, ethnicity, race, first-generation college-bound status, and other factors known to affect realization of science-related outcomes, may influence students' experiences with CUREs.

The mechanisms by which these outcomes come about and the connections between activities and outcomes and between short-, medium-, and long-term outcomes have not been empirically investigated. In addition, some outcomes demonstrated in studies of research internships are unexplored or only proposed in CUREs. For example, we found that undergraduates who complete research internships make progress in understanding of the nature of science and expand their social and professional networks (e.g., Russell *et al.*, 2007; Adedokun *et al.*, 2012; see Supplemental Table 1), but these outcomes have not been broadly explored in CUREs.

Previous work has proposed a set of features or dimensions that distinguish CUREs from other lab-learning experiences (Corwin Auchincloss *et al.*, 2014). Specifically, CUREs are thought to be distinctive because they: 1) involve students in multiple science practices, 2) provide opportunities for students to make discoveries, 3) involve students in work that has relevance outside the classroom, 4) involve students

in collaborative work, and 5) provide opportunities for iteration. The activities highlighted here relate to some but not all of these dimensions. Additional models need to be developed and tested to identify what is both necessary and sufficient in the design and implementation of CUREs to make them effective for students.

Another important avenue for study of what makes CUREs effective relates to faculty experiences and outcomes. For example, research and theories of how faculty members decide to change their teaching to more actively engage students and emphasize understanding of science concepts and practices raises interesting questions of the potential for CUREs to be a "gateway" to instructional change. For example, faculty members may be quicker to adopt evidence-based teaching approaches when they teach a CURE because of the CURE's potential to connect to their research interests and their identity as researchers (Brownell and Tanner, 2012). In addition, CUREs may resonate with faculty members' own experiences learning to do science, and thus they may be willing to sacrifice "content coverage" to offer course experiences they see as more authentic (Spell *et al.*, 2014).

CONCLUSION

In summary, we identified studies on CUREs and research internships and considered learning theory that is useful for explaining the mechanisms by which students learn through research experiences. Then, we characterized the outcomes that students have the potential to realize from participating in CUREs. Using a subset of outcomes that had the most empirical support, we used a pathway-modeling process to generate models of the components and contexts that make CUREs effective for students. We offer these models as hypotheses for the CURE community to test, revise, elaborate, or refute. We cite instruments that are ready to use in CURE assessment and note gaps for which instruments need to be developed. We hope that the community effort to assess CUREs using a common set of tools, which began with the efforts of Lopatto and colleagues (Lopatto, 2004, 2007, 2008; Lopatto *et al.*, 2008) will embrace this new set of tools such that we can continue to compare results across CUREs and between CUREs and research internships. We describe our approach so others can use it to build and test their own models of CURE instruction, which is especially important given the diverse long-term outcomes that may result from research experiences. Understanding how CUREs function will help improve existing CUREs, aid in the design of new CUREs, and promote a common understanding of the utility of CUREs as an educational intervention.

ACKNOWLEDGMENTS

Thanks to Melissa Aikens and Lucas Wachsmuth for their careful reading and thoughtful feedback on drafts of the manuscript and to the Cornell Office for Research on Evaluation for early use of the Netway (CORE, 2009) to develop pathway models. Support for this work was provided by a grant from the National Science Foundation (NSF DBI-1061874). M.J.G. is supported through a Howard Hughes Medical Institute's (HHMI) professor grant (originally to Jo Handelsman), as well as by an NSF grant (NSF-TUES 1323258). The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the HHMI or 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.
- Adedokun OA, Zhang D, Parker LC, Bessenbacher A, Chilcress A, Burgess WD (2012). Towards an understanding of the processes of the effects of undergraduate research experiences on students' aspiration for research careers and graduate education. *J Coll Sci Teach* 42, 82–91.
- Alexander BB, Foertsch J, Daffinrud S, Tapia R (1998). 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.
- Alkahr 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. Sunal D, Sunal C, Zollman D, Mason C, and Wright E, Charlotte, NC: Information Age.
- American Association for the Advancement of Science (2011). *Vision and Change in Undergraduate Biology Education: A Call to Action*, Washington, DC: <http://visionandchange.org/files/2011/03/Revised-Vision-and-Change-Final-Report.pdf> (accessed 10 January 2014).
- Bandura A (1971). *Social Learning Theory*, Englewood Cliffs, NJ: Prentice Hall.
- Barlow AE, Villarejo M (2004). Making a difference for minorities: evaluation of an educational enrichment program. *J Res Sci Teach* 41, 861–881.
- Bascom-Slack CA, Arnold AE, Strobel SA (2012). Student-directed discovery of the plant microbiome and its products. *Science* 338, 485–486.
- Beck CW, Bliwise NG (2014). Interactions are critical. *CBE Life Sci Educ* 13, 371–372.
- Bollen KA, Hoyle RH (1990). Perceived cohesion: a conceptual and empirical examination. *Social Forces* 69, 479–504.
- Boud D, Falchikov N (1989). Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *High Educ* 18, 529–549.
- Brown JS, Collins A, Duguid P (1989). Situated cognition and the culture of learning. *Educ Res* 18, 32–42.
- Brownell SE, Kloser MJ, Fukami T, Shavelson R (2012). Undergraduate biology lab courses: comparing the impact of traditionally based “cookbook” and authentic research-based courses on student lab experiences. *J Coll Sci Teach* 41, 18–27.
- Brownell SE, Kloser MJ, Fukami T, Shavelson RJ (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. *J Microbiol Biol Educ* 14, 176–182.
- Brownell SE, Tanner KD (2012). Barriers to faculty pedagogical change: lack of training, time, incentives, and tensions with professional identity. *CBE Life Sci Educ* 11, 339–346.
- Brownell SE, Wenderoth MP, Theobald R, Okoroafor N, Koval M, Freeman S, Walcher-Chevillet CL, Crowe AJ (2014). How students think about experimental design: novel conceptions revealed by in-class activities. *BioScience* 64, 125–137.
- Caruso SM, Sandoz J, Kelsey J (2009). Non-STEM undergraduates become enthusiastic phage-hunters. *CBE Life Sci Educ* 8, 278–282.
- Chemers MM, Syed M, Goza BK, Zurbriggen EL, Bearman S, Crosby FJ, Shaw JM, Hunter L, Morgan EM (2010). *The Role of Self-Efficacy and Identity in Mediating the Effects of Science Support Programs*, Technical Report no. 5, Santa Cruz: University of California.
- Chipuer HM, Pretty GMH (1999). A review of the sense of community index: current uses, factor structure, reliability, and further development. *J Commun Psychol* 27, 643–658.
- Cornell Office of Research and Evaluation (2009). *The Netway: Program and Evaluation Planning*. <https://core.human.cornell.edu/research/systems/netway.cfm> (accessed 15 September 2014).
- Corwin Auchincloss L, Laursen SL, Branchaw JL, Eagan K, Graham M, Hanauer DI, Lawrie G, McLinn CM, Pelaez N, Rowland S, et al. (2014). Assessment of course-based undergraduate research experiences: a meeting report. *CBE Life Sci Educ* 13, 29–40.
- Creamer EG, Laughlin A (2005). Self-authorship and women's career decision-making. *J Coll Stud Dev* 46, 13–27.
- Dasgupta AP, Anderson TR, Pelaez N (2014). Development and validation of a rubric for diagnosing students' experimental design knowledge and difficulties. *CBE Life Sci Educ* 13, 265–284.
- Deane T, Nomme K, Jeffery E, Pollock C, Birol G (2014). Development of the biological experimental design concept inventory (BEDCI). *CBE Life Sci Educ* 13, 540–551.
- DesJardins SL, McCall BP (2014). The impact of the Gates Millennium Scholars Program on college and post-college related choices of high ability, low-income minority students. *Econ Educ Rev* 38, 124–138.
- DesJardins SL, McCall BP, Ott M, Kim J (2010). A quasi-experimental investigation of how the Gates Millennium Scholars program is related to college students' time use and activities. *Educ Eval Policy Anal* 32, 456–475.
- Dolan EL, Lally DJ, Brooks E, Tax FE (2008). Prepping students for authentic science. *Sci Teach* 75, 38–43.
- Drew JC, Triplett EW (2008). Whole genome sequencing in the undergraduate classroom: outcomes and lessons from a pilot course. *J Microbiol Biol Educ* 9, 3–11.
- Duckworth A, Peterson C, Mathews MD, Kelly DR (2007). Grit: perseverance and passion for long-term goals. *J Pers Soc Psychol* 92, 1087–1101.
- 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, 683–713.
- Eagan MK, Newman CB (2010). *Investing in Human Capital: Underrepresented Racial Minorities' Intentions to Attend Graduate School in STEM Fields Los Angeles: UCLA Higher Education Research Institute*. <http://heri.ucla.edu/nih/downloads/AERA%202010%20-%20Eagan,%20Newman%20-%20Investing%20in%20Human%20Capital.pdf> (accessed 15 September 2014).
- 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.
- Feldon DF, Peugh J, Timmerman BE, Maher MA, Hurst M, Strickland D, Gilmore JA, Stiegemeyer C (2011). Graduate students' teaching experiences improve their methodological research skills. *Science* 333, 1037–1039.
- Glynn SM, Taasobshirazi G, Brickman P (2009). Science motivation questionnaire: construct validation with nonscience majors. *J Res Sci Teach* 46, 127–146.
- Graham MJ, Frederick J, Byars-Winston A, Bunter AB (2013). Increasing persistence of college students in STEM. *Science* 341, 1455–1456.
- Grayson J, Anderson TR, Grossley LG (2001). A four-level framework for identifying and classifying student conceptual and reasoning difficulties. *Int J Sci Educ* 23, 611–622.

- Hanauer DI, Dolan EL (2014). The project ownership survey: measuring difference 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, Ratmansky 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, Pedulla ML, Jacobs-Sera D, Cichon PM, Foley A, Ford ME, Gonda RM, Houtz JM, Hryckowian AJ, Kelchner VA, *et al.* (2006). Exploring the mycobacteriophage metaproteome: phage genomics as an educational platform. *PLoS Genetics* 2, e92.
- Hausmann LR, Schofield JW, Woods RL (2007). Sense of belonging as a predictor of intentions to persist among African American and white first-year college students. *Res High Educ* 48, 803–839.
- Hausmann LR, Ye F, Schofield JW, Woods RL (2009). Sense of belonging and persistence in white and African American first-year students. *Res High Educ* 50, 649–669.
- Heller DE (2001). *Debts and Decisions: Student Loans and Their Relationship to Graduate School and Career Choice*, Indianapolis, IN: USA Group Foundation. www.luminafoundation.org/publications/debtsdecisions.pdf (accessed 15 September 2014).
- Hunter A-B, Weston TJ, Laursen SL, Thiry H (2009). URSSA: evaluating student gains from undergraduate research in science education. *Counc Undergrad Res Q* 29, 15–19.
- Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Denney JJ, Denver DR, Dunbar D, Elgin SCR, *et al.* (2014). A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. *MBio* 5, e01051–13.
- Kloser MJ, Brownell SE, Shavelson RJ, Fukami T (2013). Effects of a research-based ecology lab course: a study of nonvolunteer achievement, self-confidence, and perception of lab course purpose. *J Coll Sci Teach* 42, 90–99.
- Laursen S, Hunter AB, 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.
- Lent RW, Brown SD, Hackett G (1994). Toward a unifying social cognitive theory of career and academic interest, choice, and performance. *J Vocat Behav* 45, 79–122.
- 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 (2008). *Classroom Undergraduate Research Experiences Survey (CURE)*. www.grinnell.edu/academics/areas/psychology/assessments/cure-survey (accessed 15 September 2014).
- Lopatto D, Alvarez C, Barnard D, Chandrasekaran C, Chung HM, Du C, Eckdahl T, Goodman AL, Hauser C, Jones CJ, *et al.* (2008). Undergraduate research: Genomics Education Partnership. *Science* 322, 684–685.
- Maton KI, Hrabowski FA, Schmitt CL (2000). African American college students excelling in the sciences: college and post-college outcomes in the Meyerhoff Scholars Program. *J Res Sci Teach* 37, 629–654.
- Mullen AL, Goyette KA, Soares JA (2003). Who goes to graduate school? Social and academic correlates of educational continuation after college. *Sociol Educ* 76, 143–169.
- Nevill SC, Chen X (2007). *The Path through Graduate School: A Longitudinal Examination 10 Years after Bachelor's Degree*, NCES-162, Washington, DC: U.S. Department of Education, National Center for Education Statistics.
- Packard BWL (2004). Mentoring and retention in college science: reflections on the sophomore year. *J Coll Stud Ret* 6, 289–300.
- 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*. www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_feb.pdf (accessed 10 September 2014).
- 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 SH, Hancock MP, McCullough J (2007). Benefits of undergraduate research experiences. *Science* 316, 548–549.
- Russell SH, 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 Prac* 12, 57–67.
- Schultz PW, Hernandez PR, Woodcock A, Estrada M, Chance RC, Aguilar M, Serpe RT (2011). Patching the pipeline: reducing educational disparities in the sciences through minority training programs. *Educ Eval Policy Anal* 33, 95–114.
- Sevian H, Gonsalves L (2008). Analysing how scientists explain their research: a rubric for measuring the effectiveness of scientific explanations. *Int J Sci Educ* 30, 1441–1467.
- Sewell WH, Haller AO, Portes A (1969). The educational and early occupational attainment process. *Am Sociol Rev* 34, 82–92.
- Seymour E, Hunter AB, 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, Alvarez C, Bailey C, Barnard D, Bhalla S, Chandrasekaran C, Chandrasekaran V, Chung H-M, Dorer DR, Du C, *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.
- Shaffer CD, Alvarez CJ, Bednarski AE, Dunbar D, Goodman AL, Reinke C, Rosenwald AG, Wolyniak MJ, Bailey C, Barnard D, *et al.* (2014). A course-based research experience: how benefits change with increased investment in instructional time. *CBE Life Sci Educ* 13, 111–130.
- Siritunga D, Montero-Rojas M, Carrero K, Toro G, Vélez A, Carro-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.
- Sirum K, Humburg J (2011). The Experimental Design Ability Test (EDAT). *Bioscene* 37, 8–16.
- Spell RM, Guinan JA, Miller KR, Beck CW (2014). Redefining authentic research experiences in introductory biology laboratories and barriers to their implementation. *CBE Life Sci Educ* 13, 102–110.
- Theobald R, Freeman S (2014). Is it the intervention of the students? Using linear regression to control for student characteristics in undergraduate STEM education research. *CBE Life Sci Educ* 13, 41–48.
- 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, 771–784.
- Thiry H, Weston TJ, Laursen SL, Hunter AB (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.
- Urban JB, Hargraves M, Trochim WM (2014). Evolutionary evaluation: implications for evaluators, researchers, practitioners,

- fundes and the evidence-based program mandate. *Eval Prog Plan* 45, 127–139.
- Urban JB, Trochim W (2009). The role of evaluation in research practice integration working toward the “golden spike.” *Am J Eval* 30, 538–553.
- Vygotsky LS (1978). *Mind in Society: The Development of Higher Mental Processes*, Cambridge, MA: Harvard University Press.
- Walpole M (2003). Socioeconomic status and college: how SES affects college experiences and outcomes. *Rev High Educ* 27, 45–73.
- Ward C, Bennett JS, Bauer KW (2002). Content analysis of undergraduate research student evaluations. www.sarc.miami.edu/ReinventionCenter/Public/assets/files/contentAnalysis.pdf (accessed 10 September 2014).
- Wenger E (1998). *Communities of Practice: Learning, Meaning, and Identity*, New York: Cambridge University Press.
- Westat JF, Mark MM, Rog DJ, Thomas V, Frierson H, Hood S, Hughes G, Johnson E (2010). *The 2010 User-Friendly Handbook for Project Evaluation*. www.purdue.edu/research/vpr/rschdev/documents/2010NSFuser-friendlyhandbookforprojectevaluation.pdf (accessed 1 September 2014).
- Wiley J (2009). *Student ownership of learning: an analysis*. Master’s thesis, Manoa: University of Hawaii.