Research Article

Measuring undergraduates' understanding of the culture of scientific research as an outcome variable in research on CUREs

Lexi Wachtell,[1] Amanda Gardiner,[1] Matt Sievers,[1] Katie Dickinson,[1] Grace E. C. Dy,[1] Elizabeth H. Glenski,[1] Joya Mukerji,[1] Elli Theobald,[1] Elisa T. Tran,[1] Vicente Velasco,[1] Scott Freeman[1]

AUTHOR AFFILIATION See affiliation list on p. 10.

ABSTRACT Researchers who work on course-based undergraduate research experiences (CUREs) and issues related to science, technology, engineering, and math (STEM) retention have begun exploring changes in student thinking about what it means to be a scientist. To support this effort, we developed rubrics to score answers to three open-response prompts: What does it mean to think like a scientist? What does it mean to do science? and Did you do real research in your *coursename* labs? The rubric development process was iterative and was based on input from the literature, experienced researchers, and early-career undergraduates. A *post hoc* analysis showed that the rubric elements map to 27 of 31 statements in the Culture of Scientific Research (CSR) framework, suggesting that scored responses to the three prompts can assess how well students understand what being a science professional entails. Scores on responses from over 400 students who were starting an introductory biology course for majors furnish baseline data from the rubrics and suggest that (i) undergraduates at this level have, as expected, a novice-level understanding of CSR, and (ii) level of understanding in novice students does not vary as a function of demography or academic preparation. Researchers and instructors are encouraged to add CSR to their list of learning objectives for CUREs and consider assessing it using the rubrics provided here.

KEYWORDS undergraduate STEM education, undergraduate labs, CUREs, thinking like a scientist, doing science, real research

Course-based undergraduate research experiences (CUREs) focus on authentic research—questions where the answers are unknown and the data are potentially publishable (1–4). CUREs aim to broaden access to the well-documented benefits of classical, apprentice-style undergraduate research experiences (UREs)—primarily increased retention in science, technology, engineering, and math (STEM) majors and stronger commitments to pursuing STEM-related careers (5–9). Recent studies link CUREs to constructs that predict retention in STEM, including increased identity as a scientist, sense of belonging in science, science self-efficacy, interest in pursuing a URE, and stated career intentions (10), as well as gains in more direct measures of STEM persistence and retention such as graduation rates (11).

To complement research on the overall goal of retention, Brownell and co-workers have been exploring how CURE labs impact a more specific learning outcome. They asked participants to respond to the prompts, "What does it mean to think like a scientist?" (12) and "Did you participate in real research?" (3). Their intent was to evaluate how well students understand the nature of scientific investigation (sensu 13)—meaning, "how scientists do their work and how … scientific knowledge is generated and accepted" (14, p. 66). This construct has recently been expanded into a more granular and testable framework called the culture of scientific research, or CSR (15). CSR is focused on what it means to be a scientist and is related to but distinct

Editor Lisa K. Elfring, The University of Arizona, Tucson, Arizona, USA

Address correspondence to Scott Freeman, srf991@uw.edu.

Lexi Wachtell, Amanda Gardiner, and Matt Sievers are joint first authors. Order of authorship was determined by random draw.

The authors declare no conflict of interest.

Received 7 October 2022
Accepted 18 July 2023
Published 15 September 2023

from well-established frameworks for understanding the nature of science (NOS). CSR emphasizes science as a way of being and doing—of conducting oneself as a professional—while NOS emphasizes science as a way of knowing, including how science compares to religion and to other modes of inquiry and understanding (14).

In essence, the learning goal that Brownell and co-workers set before the CURE community is to measure whether students understand how scientists think, what they value, and what they do in terms of day-to-day practice (15). As Dewey and colleagues (15) point out, entering the realm of science in this way represents what scholars call a cultural, emotional, and professional "border crossing"—one that can be particularly difficult for students from groups that are minoritized in STEM (16).

For decades, researchers have documented naïve ideas about CSR-like constructs by asking K-16 students to "draw a scientist" (17). A recent meta-analysis confirmed that most responses depict a white male who works in a lab surrounded by glassware and solutions, although this male-bias has declined over time (18). Stereotypical views like these—developed from images on television, in movies, books, and social media—can negatively impact students' sense of belonging in science and success in undergraduate STEM courses (19, 20).

Following Brownell et al. (12), we propose that CURE developers and researchers expand their learning goals and assessments to explore how well students understand CSR. The question is consistent with the original motivation behind CUREs because elements of culture—specifically identifying as a scientist and embracing the values of science—are integral to the success of UREs for minoritized groups in STEM (8).

Early work by Brownell's group has shown promising gains in elements of CSR, suggesting that pursuing the construct is both relevant and achievable. Specifically, Brownell et al. (12) pre-post tested students in an upper-division CURE with the prompt, "What do you think it means to think like a scientist?" The team coded the responses into themes and categories and observed CURE participants to become more expert-like, based on criteria in the literature. Subsequently, Cooper et al. (3) asked upper-division students in closely aligned traditional and CURE labs whether they had done "real research," with a 10-point Likert scale response followed by a prompt to explain their answer. Students in the CURE lab ranked the experience as "more real," with their explanations emphasizing that the data they produced were novel and relevant to the scientific community.

More recently, an interview study coded student responses to three broad questions reflecting on their experiences in a CURE, with codes corresponding to each element in the CSR (21). The results showed striking differences in which elements of the CSR were emphasized, based on whether students participated in bench-based or computer-based projects. These results suggest that fundamental aspects of a CURE's design can impact how students perceive the culture of scientific research and should inspire further work exploring the role of CUREs in advancing understanding of scientific culture.

The goal of this study is to support CURE researchers in the ongoing effort to quantify students' understanding of CSR. Specifically, we offer rubrics to quantify students' answers to the two prompts that have been used in previous work on CSR, along with a related third prompt:

1. What does it mean to think like a scientist?
2. What does it mean to do science?
3. Did you perform what you would call real research in your *coursename* labs? Why or why not?

Open-response prompts like this, though more difficult to grade, can produce a more robust understanding of student thinking than fixed-response item formats (22).

After developing the rubrics, we used them to score responses from students who were just starting an introductory biology series for life sciences majors. In addition to offering preliminary data on student understanding of CSR, the scoring process allowed us to assess how feasible it will be for researchers to hand-score responses in large-scale experiments on CUREs.

METHODS

Rubric development

Brownell and co-workers coded student answers to the "think like a scientist" and real research prompts in a bottom-up approach, then characterized responses as naïve or expert-like based on the literature. To complement and extend this progress, we pursued a top-down strategy based on literature review and expert responses to the think like a scientist, "doing science," and real research prompts. Our goal was to support content validity with an aspirational rubric that would reflect as many aspects of expert thinking as possible (23). We then tested and revised this expert-level framework against actual student responses—prior to finalizing the rubric and administering it at scale—in an iterative fashion.

For each prompt, rubric development began with a literature search to identify key aspects of scientific thinking and practice (Fig. 1). For example, recent work characterizes scientific thinking as making discoveries (12, 24–27), making connections between seemingly unconnected phenomena (12, 24), critically evaluating data with skepticism (12, 23, 24), and seeking opportunities to share findings and communicate with others (25).

The literature was also helpful in drafting the rubric on real research. We relied heavily on groundbreaking work on the question of "what makes a CURE a CURE" (2), the Laboratory Course Assessment Survey (LCAS; 28), and a preparation course for students interested in UREs (28). These studies identified four overarching themes in characterizing authentic research: (i) motivation or goal, (ii) process, (iii) iteration, and (iv) sense of project ownership. They also specified key elements within each theme. To create an initial draft rubric based on these themes, we used elements in Table 2 of Auchincloss et al. (27), items in the LCAS (29), and aspects of project ownership emphasized by Cartrette and Miller (28).

We built on the literature review by analyzing a minimum of 10 expert responses to each prompt. The experts were blind to the nature of the study and were primarily life sciences research faculty, although several were graduate students or advanced undergraduates with extensive research experience. The faculty experts were chosen because they not only were mentoring postdocs, graduate students, and undergraduates in their research labs but also had expressed a commitment to teaching aspects of research culture in the upper-division courses they taught. Likewise, the graduate students involved were actively mentoring undergraduate researchers in experimental design and the scientific process. We included two advanced undergraduates under the hypothesis that they might articulate insights into scientific culture that might be taken for granted by more experienced researchers but which for them were recently discovered or personally surprising or impactful. Each expert response was coded by at least two experienced undergraduate researchers who were blind to the goal of the study. After identifying key elements independently, coders met to reach consensus on themes that were declared by more than one expert. If they were not already among the themes identified from the literature, these elements were added to create a revised draft of the rubric for each prompt (Fig. 1).

We evaluated these second draft rubrics based on responses to the prompts provided by 100 students taking an introductory biology course for majors. We analyzed responses from the think like a scientist and doing science recorded at the start of the course and from the real research prompt—which is retrospective—administered at the end of the course. To enrich our analysis of the real science prompt, we arranged the sample such that half of the responses came from students who were doing expository or inquiry labs and half from students who were doing a CURE (30). Two members of the research team

FIG 1 Steps in rubric development. All three prompts followed the development sequence summarized here.

scored each response independently using the draft rubrics, then met to reach consensus. This step in our development sequence (Fig. 1) had two goals:

- Annotating the rubric to begin standardizing scoring decisions. For example, coders confirmed with the rest of the research team that they would not mark positive scores if responses used buzzwords or pat phrases such as "use the scientific method," instead of more specific language that both conformed to elements in the rubric and clearly indicated an expert-like understanding. In addition, the real science rubric was modified to accommodate "yes" and "no" responses, so that students would receive points for giving an expert-like reason why their lab was not real research.
- Identifying common elements in the responses that were missing in the draft rubrics. This step was designed to increase face validity by ensuring that the final versions captured as full a range of student responses as possible (22). For example, students who claimed that they were doing real research in their course labs frequently mentioned that the work was long term and that it was part of a larger and ongoing investigation, which in their minds appeared important in distinguishing the CURE from traditional "one-off" labs. If student responses such as these were validated by discussion among members of the research team, we returned to the literature to assess support and refined the rubric accordingly.

Unfortunately, rubrics scored on a 0/1 system are not amenable to exploratory factor analysis, which is the most productive way to evaluate construct validity (31). We were able to assess construct validity in a preliminary way, however, using the CSR framework (15), which was published after all three of our rubrics were finalized. *Post hoc*, two members of the research team (K.D. and S.F.) independently matched each element in the think like a scientist, do science, and real research rubrics to a statement in the CSR framework, then met to reach consensus on which CSR components were represented by specific elements in the rubrics.

Data collection and analysis

To test how well the rubrics work when used at scale and to document results in a population of prospective STEM majors, we administered the prompts to all students at the start of a large-enrollment introductory biology course for majors, over two terms. These students were taking traditional labs but were studied to provide (i) baseline data for this paper and (ii) a comparison group for a planned CURE intervention. By timing the responses at the start of the term, we hoped to quantify the views of first- and second-year undergraduates before they had completed all of their core introductory STEM courses or declared a major, and when they were unlikely to have completed a classical, apprentice-style undergraduate research experience.

Responses for each rubric were scored by two advanced undergraduates who were blind to the goal of the study. After an initial training session with sample student responses that was facilitated by a member of the research team, the raters followed an iterative process of grading identical questions independently and meeting to reach consensus, until inter-rater reliability scores exceeded 0.80. Once that threshold was exceeded, each rater scored student responses independently.

To assess which elements of the rubrics were less frequently articulated by novice learners than others, we calculated the raw percent correct—with "correct" indicating expert-like—for each element in each rubric. To test the hypothesis that summed scores at the start of instruction varied as a function of student demographic characteristics or academic preparation and ability—specifically that students from minoritized groups differ from other students in their views of CSR—we used data from the university's Registrar to run binomial regression models with sex, race/ethnicity, first-generation status, socio-economic status (SES), and total SAT score as predictors. We used binomial regressions as they are appropriate to data that are expressed as proportions (32)—in this case, the percent of expert-like responses to a given prompt.

RESULTS

The rubrics

The final rubrics are summarized in Table 1. Note that there are six broad themes in the thinking like a scientist rubric (Table 1A). As a result, student understanding of scientific thinking could be measured as the simple sum of the 0/1 scores for each theme, meaning that a team of experts would be expected to produce an answer with a total score of 6.

In the case of the doing science rubric, however, the elements fell naturally into five categories, each with two to four sub-elements (Table 1B). Although some of the sub-elements within categories appear to be ordinal—meaning that they may represent an increasingly sophisticated understanding of what it means to do a particular aspect of science—we recommend that scores across the rubric simply be summed to create an index of whether students have an expert-like understanding of how science is done.

We also interpreted the elements in the real research rubric as falling into five categories. In this case, we interpreted the sub-elements in each of the five categories as equally valid and non-overlapping ways to express each theme (Table 1C). In addition, responses start with a yes or no followed by an explanation, so researchers have several options for quantifying student understanding. These options include the following: (i) the percentage of the five categories with a "hit" after a "yes" answer, an approach that would gauge whether students thought they had experienced the major elements in authenticity; (ii) the simple sum of points across all five categories after a "yes" answer, which would provide an index of the overall experience of authenticity, (iii) the probability of answering yes or no; and/or (iv) the probability of giving expert-like warrants to explain a "no" answer.

Total training time for the assistants who scored responses was approximately 6 hours for the three rubrics, in total. Once training resulted in inter-rater reliabilities of 0.80 or higher on each rubric, assistants were able to score a student response in less than 2 minutes, on average. This means that a single undergraduate researcher could reliably evaluate over a hundred responses to one of the prompts in an hour. Copies of the full rubrics with sample student responses and scores are provided in Supplemental Material, Appendix 1.

The numbers in the right-most column of Table 1A indicate the percent correct for each rubric element in responses to the thinking like a scientist prompt. At this point in their careers, about a third of the students in our sample declares that scientists ask questions and design experiments to answer them, and about one in five characterizes critical thinking and open-mindedness as key aspects of thinking like a scientist. But almost none stated that scientists make decisions based on evidence or that conclusions in science are strengthened by evidence from multiple sources.

The data on percent correct for the doing science prompt, in the right-most column of Table 1B, indicate that about 60% of surveyed students consider doing experiments fundamental, while 30%–40% stated that scientists make observations and ask questions. Between 10% and 20% referenced hypothesis testing and data collection, but less than 1 student in 20 mentioned the other 10 sub-elements in the rubric.

TABLE 1 The final rubrics

Category	Explanation/examples (rubric A) or sub-element (rubrics B and C)	% Responses including this element
A. What does it mean to think like a scientist? (*n* = 424)		
Asking questions	Curiosity, extend frontier of knowledge	36.1
Process thinking	Hypothesis testing, experimental design	31.6
Critical thinking	Skepticism, demanding evidence, quality assurance, rigor	13.9
Evidence-based conclusions	Data-based reasoning	4.2
Open minded	Consider alternatives, multiple perspectives	16.7
Multiple approaches	Most convincing evidence is based on multiple independent sources	0.5
B. What does it mean to do science? (*n* = 415)		
Investigate	Consult prior studies	4.1
	Observe natural world	41.0
	Ask a question	30.8
Collect data	Perform an experiment or collect observational data	61.4
	Test a hypothesis	18.6
	Repeat the experiment to verify the result	2.6
Analyze data	Analyze data (include visualization)	6.5
	Interpret data	11.1
	Patterns may lead to models	0.0
Collaborate	Work in a team	0.5
	Exchange information and ideas among team members	2.4
	Jointly produce information for dissemination	0.0
Communicate	Share results with community (papers, posters, etc.)	3.1
	Undergo peer review	0.5
	Replicate other teams' findings	1.7
C. Did you do real research in your *coursename* labs?		
Authenticity	New knowledge	
	Relevance to scientific community	
Processes	Collaboration	
	Used publication-standard techniques	
	Understand how and why the techniques work	
	No right/wrong data	
Iteration	Troubleshoot	
	Repeat experiments	
Connections to other work	Work continued over course of term	
	Work will continue beyond the class	
	Communicate results	
Ownership	Work on own question and/or hypothesis	
	Design the experiment	

(*Continued on next page*)

TABLE 1 The final rubrics (*Continued*)

Category	Explanation/examples (rubric A) or sub-element (rubrics B and C)	% Responses including this element
	Carry out the experiment or observations	
	Be responsible for the integrity of the data	

Figure 2 provides the distributions of summed pre-scores on the thinking like a scientist and doing science prompts. Model output from the analyses of pre-scores on these two prompts is given in Appendix 2. The regression results indicate that none of the demographic variables in the model explained a significant amount of variation in the summed scores on these two prompts. As an index of academic preparation and ability, SAT total score was also not an important predictor.

Table 2 shows how elements in the three rubrics mapped onto the CSR framework of Dewey et al. (15). Of the 31 statements in the framework, 27 correspond to at least one element in the three rubrics according to independent evaluations by two members of the research team. The CSR statements not assessed by the three prompts and rubrics in this study are the following: scientists should have freedom and independence; science is not all-knowing; science is influenced by and contributes to society and culture; and science is constructive and complex.

DISCUSSION

The thinking like a scientist, doing science, and real research prompts and rubrics reported here assess student understanding of 87% of the Culture of Scientific Practice framework statements. This result suggests that in combination, the three questions and scoring rubrics offer a valuable tool for researchers who want to explore the development of student thinking about what it means to be a scientist in terms of values and practice. In addition, the short training time, ability to engage advanced undergraduates as coders and achieve high inter-rater reliability, and the brief time to mark each question accurately all suggest that these three rubrics are practical to use in large-scale studies.

It is important to note, however, that the raters who evaluated the correspondence between the rubric elements and CSR identified three CSR statements that are not completely evaluated in the rubrics, in addition to the four statements that are not covered at all. The "partial-coverage" statements are (i) scientists aim to be objective but are influenced by their prior knowledge and beliefs, (ii) scientists must publish their work as a measure of success, often leading to competition, and (iii) scientists must be open to new ideas but can be influenced by personal bias. In each, there is at least one element in the rubrics that conforms to the initial part of each statement, but none that reference the second parts—i.e., "are influenced by their prior knowledge and beliefs," "often leading to competition," and "can be influenced by personal bias." In addition, researchers who are interested in using the prompts and rubrics to quantify student understanding of CSR should note that the think like a scientist and doing science rubrics cover 26 of the 27 CSR elements—meaning that including the real research prompt may be more relevant to other study objectives.

The distributions of summed scores (Fig. 2) suggest that students who are just starting an introductory series for life sciences majors have far from an expert-like understanding of CSR. In addition, the model output (Appendix 2) indicates that scores did not vary as a function of student characteristics. Scores were extremely low for all students in this sample, with few indicating that the culture of scientific research involves more than asking questions and doing experiments. These results are not surprising, given that almost all of the students were novices in terms of their direct experience with scientific research. The baseline data reported here suggest that interventions like CUREs

TABLE 2 Mapping rubric elements to statements in the CSR framework of Dewey et al. (15)

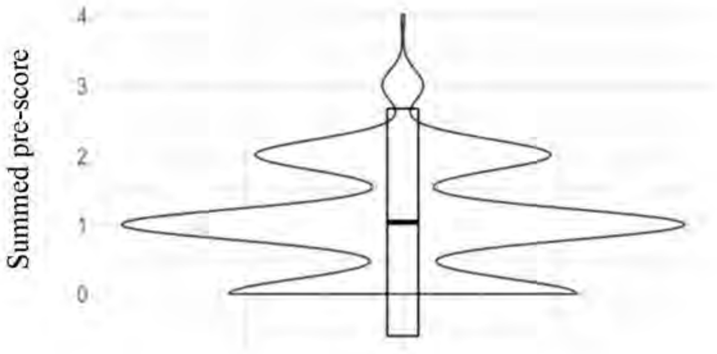
CSR category and statement		"Thinking like a scientist" category	"Doing science" category (and statement)	Real research category (and statement)
Practices	Pose questions, hypotheses, predictions	Ask questions; process thinking	Investigate (develop question)	Ownership (work on own question/hypothesis)
	Plan investigation	Process thinking	Gather data (test hypotheses)	Ownership (design experiment)
	Run investigation	Process thinking	Gather data (perform experiment)	Ownership (carry out experiment)
	Analyze data		Analyze data (… includes graphing)	
	Interpret data	Evidence-based conclusions	Analyze data (interpret data)	Process (no right/wrong)
	Generate explanations and conclusions		Analyze data (interpret data)	
	Negotiate and debate		Collaboration (exchange info/ feedback)	
	Produce and use representations		Analyze data (… includes graphing)	
	Develop and use models		Analyze data (patterns to models)	
	Use quantitative approaches		Analyze data (… includes graphing)	
	Obtain and evaluate information		Gather data (perform experiment); analyze data (interpret data)	
	Communicate		Communicate (share deliverables)	Connect to other work (communicate results)
	Teamwork		Collaboration (all three statements)	Process (collaboration)
Norms/expectations	Aim to be objective	Critical thinking		
	Aim for integrity	Critical thinking		Ownership (data integrity)
	Work should be repeated		Communicate (replicate and verify others); gather data (repeat experiment)	Iteration (repeat experiments)
	Work is often peer reviewed		Communicate (peer review)	
	Publish as a measure of success		Communicate (share deliverables)	
	Often collaborative		Collaboration (all three statements)	Process (collaboration)
	Freedom and independence			
	Persistence and resilience			Iteration (troubleshoot)
	Open to new ideas	Multiple perspectives		
Values/beliefs	Discover new knowledge	Ask questions	Investigate (all three statements)	Authenticity (new knowledge)
	Requires empirical evidence	Evidence-based conclusions; critical thinking		
	Not all-knowing			
	Produces durable but tentative knowledge	Critical thinking		
	Importance of curiosity, imagination, creativity	Ask questions		
	Not defined by a single method	Multiple approaches		
	Influenced by society and contributes to it			
	Builds on what has gone before	Multiple approaches	Investigate (consult prior studies); communicate (verify others')	
	Constructive and complex			

may have the potential to support important changes in how well students understand what it means to be a scientist.

Future work

We view the prompts and rubrics offered here as a work in progress. Future research may show that altered or additional prompts are needed to assess student understanding of the four statements that are not included in the rubrics and/or the three statements

A. “Thinking like a scientist” prompt (6 points possible; $n = 424$)



B. “Doing science” prompt (15 points possible; $n = 415$)

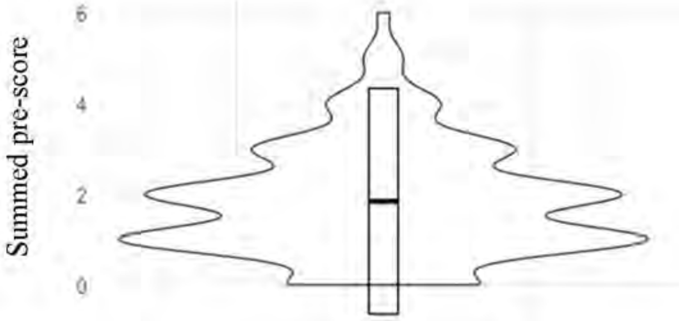


FIG 2 Distributions of summed pre-instruction scores for the “Thinking like a scientist” prompt (A) and the “Doing science” prompt (B). These are boxplots of summed scores superimposed on Violin plots. Violin plots show the complete data in kernel density plots (smoothed histograms) along the vertical axis, presented symmetrically to support easier interpretation.

that are only partially covered. Expanding or changing prompts has been a productive practice with other open-response assessments, leading to richer or more interpretable student responses (33, 34).

Future work on construct validity might consist of comparing scoring on open responses to the think like a scientist, doing science, and real science prompts to data from related, scaled surveys such as the LCAS (29) and the project ownership survey (35). It may also be possible to probe construct validity further using think-alouds or other interview techniques to explore alternative ways that students express key ideas represented in the rubrics and to test our assumption that the use of buzzwords or pat phrases does indeed mask naïve ideas or indicate gaps in understanding.

Even if future work shows that the prompts and rubrics developed and evaluated here can be improved, recent work has shown that they can be useful in their present form. Freeman et al. (36) incorporated the prompts as an outcome variable in a study of how a CURE on the evolution of antibiotic resistance impacted students compared to traditional labs. Using the rubrics and the same iterative coder-training protocol reported here, they showed that CURE students made significantly higher gains in understanding CSR than did students in the traditional labs. Although this result needs to be replicated at additional institutions and with other CUREs before researchers can be confident that CUREs are more effective at supporting understanding of CSR than traditional labs, it is consistent with other recent research. For example, Gin et al. (37) found that student experiences of research culture during the same CURE can vary, in their case depending on the degree of difficulty encountered with samples or protocols. Similarly, Dewey et al. (20) showed that when students were interviewed about their research experiences, they mentioned different aspects of the CSR based on whether they had done a bench- or computer-based CURE. In addition to exploring whether CUREs lead to improved outcomes compared to classroom-only courses, traditional labs, inquiry labs, or UREs, it would be interesting to know whether certain CUREs lead to better progress on the Culture of Scientific Research construct compared to other CUREs.

We also urge CURE researchers to consider studies on additional questions, such as:

- Do structural equation models or other approaches support the hypothesis that a more expert-like understanding of CSR leads to increased identity as a scientist, science self-efficacy, course performance, willingness to do a URE, and/or intent to pursue a STEM-related major or career? Stated another way, does understanding what it means to be a scientist make a student more likely to border-cross and embrace the culture of science as their own? If so, then understanding CSR may be integral developing a science self, sensu Schinske and co-workers (19). Examining these questions could be particularly valuable if researchers have access to data that are disaggregated, allowing them to test the hypothesis that when students abandon a “white-guy-in-a-labcoat” stereotype in favor of more sophisticated views of science culture, disproportionate benefits accrue for students from groups that are minoritized in STEM.
- Can longitudinal studies, based on repeated use of the prompts and rubrics presented here, identify patterns in how students transition from naïve ideas about what it means to be a scientist to a more expert-like understanding over the course of an undergraduate career? Stated another way, can future research identify a robust learning progression from novice-expert level of understanding the CSR?

As the Buchanan and Fisher (4) review shows, the published literature on CUREs is growing rapidly and is reflecting an increasingly sophisticated program of research on student outcomes—including on the question, “Do CUREs help students learn how to think like scientists?”

ACKNOWLEDGMENTS

We thank Grace Dy, Elizabeth Glenski, and Elisa Tran for assistance with analyzing expert responses to the three prompts used in this study, and Margot Maraghe, Aiva Petriceks, and Louise Sutters for scoring student responses to the prompts analyzed here.

This work was made possible by grant number 52008126 from the Howard Hughes Medical Institute and was conducted under University of Washington Human Subjects Division application 00003631.

AUTHOR AFFILIATION

[1]Department of Biology, University of Washington, Seattle, Washington, USA

PRESENT ADDRESS

Joya Mukerji, Department of Biological Sciences, Sacramento State University, Sacramento, California, USA

AUTHOR ORCIDs

Scott Freeman http://orcid.org/0000-0003-0988-1900

AUTHOR CONTRIBUTIONS

Lexi Wachtell, Conceptualization, Data curation, Investigation, Methodology, Project administration, Writing – review and editing | Amanda Gardiner, Conceptualization, Data curation, Writing – review and editing, Investigation, Methodology, Writing – review and editing | Matt Sievers, Conceptualization, Data curation, Writing – review and editing, Investigation, Methodology, Project administration, Writing – review and editing | Katie Dickinson, Conceptualization, Data curation, Project administration, Writing – review and editing | Grace E. C. Dy, Conceptualization, Data curation, Investigation, Writing – review and editing | Elizabeth H. Glenski, Conceptualization, Data curation, Writing – review and editing, Investigation, Methodology, Writing – review and editing | Joya Mukerji, Conceptualization, Methodology, Writing – review and editing | Elli Theobald, Writing – review and editing, Methodology, Writing – review and editing | Elisa T. Tran, Conceptualization, Data curation, Investigation, Writing – review and editing | Vicente Velasco, Writing – review and editing, Methodology, Writing – review and editing | Scott Freeman, Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Writing – original draft, Writing – review and editing

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Appendix 1 and Appendix 2 (jmbe00187-22-S0001.pdf). Sample student responses to prompts and regression output.

REFERENCES

1. Brownell SE, Kloser M, 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 41:36–45.
2. Auchincloss LC, Laursen SL, Branchaw JL, Eagan K, Graham M, Hanauer DI, Lawrie G, McLinn CM, Pelaez N, Rowland S, Towns M, Trautmann NM, Varma-Nelson P, Weston TJ, Dolan EL. 2014. Assessment of course-based undergraduate research experiences: a meeting report. CBE Life Sci Educ 13:29–40. https://doi.org/10.1187/cbe.14-01-0004
3. Cooper KM, Blattman JN, Hendrix T, Brownell SE. 2019. The impact of broadly relevant novel discoveries on student project ownership in a traditional lab course turned CURE. CBE Life Sci Educ 18:ar57. https://doi.org/10.1187/cbe.19-06-0113
4. Buchanan AJ, Fisher GR. 2022. Current status and implementation of science practices in course-based undergraduate research experiences (CUREs): a systematic literature review. CBE Life Sci Educ 21:ar83. https://doi.org/10.1187/cbe.22-04-0069
5. Wei CA, Woodin T. 2011. Undergraduate research experiences in biology: alternatives to the apprenticeship model. CBE Life Sci Educ 10:123–131. https://doi.org/10.1187/cbe.11-03-0028
6. Bangera G, Brownell SE. 2014. Course-based undergraduate research experiences can make scientific research more inclusive. CBE Life Sci Educ 13:602–606. https://doi.org/10.1187/cbe.14-06-0099
7. Gentile J, Brenner K, Stephens A, Committee on strengthening research experiences for undergraduate STEM students, board on science education, division of behavioral and social sciences and education, board on life sciences, division on earth and life studies, national academies of sciences, engineering, and medicine. 2017. Undergraduate research experiences for undergraduates: successes, challenges, and opportunities. National Academies Press, Washington, D.C. https://doi.org/10.17226/24622
8. Estrada M, Hernandez PR, Schultz PW. 2018. A longitudinal study of how quality mentorship and research opportunities integrate underrepresented minorities into STEM careers. CBE Life Sci Educ 17:ar9. https://doi.org/10.1187/cbe.17-04-0066
9. Wilson AE, Pollock JL, Billick I, Domingo C, Fernandez-Figueroa EG, Nagy ES, Steury TD, Summers A. 2018. Assessing science training programs: structured undergraduate research programs make a difference. Biosci 68:529–534. https://doi.org/10.1093/biosci/biy052
10. Corwin LA, Runyon CR, Ghanem E, Sandy MClarkGPalmer GC, Reichler S, Rodenbusch SE, Dolan EL. 2018. Effects of discovery, iteration, and collaboration in laboratory courses on undergraduates' research career intentions fully mediated by student ownership. CBE-Life Sci Educ 17:ar20. https://doi.org/10.1187/cbe.17-07-0141
11. Rodenbusch SE, Hernandez PR, Simmons SL, Dolan EL. 2016. Early engagement in course-based research increases graduation rates and completion of science, engineering, and mathematics degrees. CBE Life Sci Educ 15:ar20. https://doi.org/10.1187/cbe.16-03-0117

12. Brownell SE, Hekmat-Scafe DS, Singla V, Chandler Seawell P, Conklin Imam JF, Eddy SL, Stearns T, Cyert MS. 2015. A high-enrollment course-based undergraduate research experience improves student conceptions of scientific thinking and ability to interpret data. CBE Life Sci Educ 14:ar21. https://doi.org/10.1187/cbe.14-05-0092
13. Nehring A. 2020. Naïve and informed views on the nature of scientific inquiry in large-scale assessments: two sides of the same coin or different currencies?. J Res Sci Teach 57:510–535. https://doi.org/10.1002/tea.21598
14. Lederman JS, Lederman NG, Bartos SA, Bartels SL, Meyer AA, Schwartz RS. 2014. Meaningful assessment of learners' understandings about scientific inquiry—the views about scientific inquiry (VASI) questionnaire. J Res Sci Teach 51:65–83. https://doi.org/10.1002/tea.21125
15. Dewey J, Roehrig G, Schuchardt A. 2021. Development of a framework for the culture of scientific research. CBE Life Sci Educ 20:ar65. https://doi.org/10.1187/cbe.21-02-0029
16. Aikenhead GS. 1996. Science education: border crossing into the subculture of science. Stud Sci Educ 27:1–52. https://doi.org/10.1080/03057269608560077
17. Farland-Smith D. 2012. Development and field test of the modified draw-a-scientist test and the draw-a-scientist rubric. Sch Sci Math 112:109–116. https://doi.org/10.1111/j.1949-8594.2011.00124.x
18. Miller DI, Nolla KM, Eagly AH, Uttal DH. 2018. The development of childrens' gender-science stereotypes: a meta-analysis of 5 decades of U.S. draw-a-scientist studies. Child Dev 89:1943–1955. https://doi.org/10.1111/cdev.13039
19. Cheryan S, Master A, Meltzoff AN. 2015. Cultural stereotypes as gatekeepers: increasing girls' interest in computer science and engineering by diversifying stereotypes. Front Psychol 6:ar49. https://doi.org/10.3389/fpsyg.2015.00049
20. Schinske JN, Perkins H, Snyder A, Wyer M. 2016. Scientist spotlight homework assignments shift students' stereotypes of scientists and enhance science identity in a diverse introductory science class. CBE Life Sci Educ 15:ar47. https://doi.org/10.1187/cbe.16-01-0002
21. Dewey J, Evers A, Schuchardt A. 2022. Students' experiences and perceptions of the scientific research culture after participating in different course-based undergraduate research experience models. CBE Life Sci Educ 21:ar36. https://doi.org/10.1187/cbe.21-10-0304
22. Hubbard JK, Potts MA, Couch BA. 2017. How question types reveal student thinking: an experimental comparison of multiple true-false and free-response formats. CBE Life Sci Educ 16:ar26. https://doi.org/10.1187/cbe.16-12-0339
23. Anastasi A, Urbina S. 1997. Psychological testing. 7th ed. Prentice Hall, Upper Saddle River NJ.
24. Druger M, Siebert ED, Crow LW, eds. 2004. Teaching tips: innovations in undergraduate science instruction. NSTA Press, Arlington VA.
25. Hunter A-B, Laursen SL, Seymour E. 2007. Becoming a scientist: the role of undergraduate research in students' cognitive, personal, and professional development. Sci Educ 91:36–74. https://doi.org/10.1002/sce.20173
26. Hurtado S, Eagan K, Sharkness J. 2009. Thinking and acting like a scientist: investigating the outcomes of introductory science and math courses. Association of Institutional Researchers Forum, Atlanta, GA.
27. Etkina E, Planinšič G. 2014. Thinking like a scientist. Phys World 27:48–51. https://doi.org/10.1088/2058-7058/27/03/40
28. Cartrette DP, Miller ML. 2013. Purposeful design of formal laboratory instruction as a springboard to research participation. J Chem Educ 90:171–177. https://doi.org/10.1021/ed2006983
29. Corwin LA, Runyon C, Robinson A, Dolan EL. 2015. The laboratory course assessment survey: a tool to measure three dimensions of research-course design. CBE Life Sci Educ 14:ar37. https://doi.org/10.1187/cbe.15-03-0073
30. Dickinson K, Mukerji J, Graham S, Warfield L, Kerr B. 2021. A course-based undergraduate research experience on the evolution of antibiotic resistance and its molecular basis. Preprints. https://doi.org/10.20944/preprints202107.0420.v1
31. Knekta E, Runyon C, Eddy S, Brickman P. 2019. One size doesn't fit all: using factor analysis to gather validity evidence when using surveys in your research. CBE Life Sci Educ 18:rm1. https://doi.org/10.1187/cbe.18-04-0064
32. Theobald EJ, Aikens M, Eddy S, Jordt H. 2019. Beyond linear regression: a reference for analyzing common data types in discipline based education research. Phys Rev Phys Educ Res 15:020110. https://doi.org/10.1103/PhysRevPhysEducRes.15.020110
33. 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. https://doi.org/10.1093/biosci/bit016
34. Sievers M, Reemts C, Dickinson KJ, Mukerji J, Beltran IB, Theobald EJ, Velasco V, Freeman S. 2023. Assessing how well students understand the molecular basis of evolution by natural selection. Biochem Mol Biol Educ 51:212–220. https://doi.org/10.1002/bmb.21697
35. Hanauer DI, Dolan EL, Smith M. 2014. The project ownership survey: measuring differences in scientific inquiry experiences. CBE Life Sci Educ 13:149–158. https://doi.org/10.1187/cbe.13-06-0123
36. Freeman S, Mukerji J, Sievers M, Beltran IB, Dickinson K, Dy GEC, Gardiner A, Glenski EH, Hill MJ, Kerr B, Monet D, Reemts C, Theobald E, Tran ET, Velasco V, Wachtell L, Warfield L. 2023. A CURE on the evolution of antibiotic resistance in *Escherichia coli* improves student conceptual understanding. CBE Life Sci Educ 22:ar7. https://doi.org/10.1187/cbe.21-12-0331
37. Gin LE, Rowland AA, Steinwand B, Bruno J, Corwin LA. 2018. Students who fail to achieve predefined research goals may still experience many positive outcomes as a result of CURE participation. CBE Life Sci Educ 17:ar57. https://doi.org/10.1187/cbe.18-03-0036