AMERICAN SOCIETY FOR MICROBIOLOGY | JMBE Journal of Microbiology & Biology Education

| Research Article

A framework for leveraging network course-based undergraduate research experience (CURE) faculty to develop, validate, and administer an assessment instrument

Adam J. Kleinschmit,[1] Elizabeth Genné-Bacon,[2] Kevin Drace,[3] Brinda Govindan,[4] Jennifer R. Larson,[5] Amber A. Qureshi,[6] Carol Bascom-Slack[2]

AUTHOR AFFILIATIONS See affiliation list on p. 15.

ABSTRACT Over the last several years, nationally disseminated course-based undergraduate research experiences (CUREs) have emerged as an alternative to developing a novel CURE from scratch, but objective assessment of these multi-institution (network) CUREs across institutions is challenging due to differences in student populations, instructors, and fidelity of implementation. The time, money, and skills required to develop and validate a CURE-specific assessment instrument can be prohibitive. Here, we describe a co-design process for assessing a network CURE [the Prevalence of Antibiotic Resistance in the Environment (PARE)] that did not require support through external funding, was a relatively low time commitment for participating instructors, and resulted in a validated instrument that is usable across diverse PARE network institution types and implementation styles. Data collection efforts have involved over two dozen unique institutions, 42 course offerings, and over 1,300 pre-/post-matched assessment record data points. We demonstrated significant student learning gains but with small effect size in both content and science process skills after participation in the two laboratory sessions associated with the core PARE module. These results show promise for the efficacy of short-duration CUREs, an educational research area ripe for further investigation, and may support efforts to lower barriers for instructor adoption by leveraging a CURE network for developing and validating assessment tools.

KEYWORDS prevalence of antibiotic resistance in the environment (PARE), course-based undergraduate research experience (CURE), course-based research experience (CRE), assessment development and validation, learning gains, science process skills, multi-institution short-duration modular network CURE, antimicrobial-resistant bacteria, science education

Engaging undergraduates in research has been identified as a high-impact educational practice (1). However, lack of access to research opportunities and lack of participation of underrepresented groups have been identified as barriers to the equitable implementation of this high-impact practice (2, 3). Engagement in authentic research, exploring the scientific process, practicing collaboration, and integrating scientific discovery within a required course are a pedagogical opportunity to address this issue in the form of a course-based undergraduate research experience (CURE) (4–9). In 2012, the Course-Based Undergraduate Research Experiences Network was established to promote the integration of research experiences into undergraduate courses to enhance student learning as well as understand and identify the direction for CURE assessments and evaluations (10).

CURE assessment has evolved in the last several years from student self-reports of outcomes from individual CURE participation to more sophisticated modeling and

Editor Min-Ken Liao, Furman University, Greenville, South Carolina, USA

Address correspondence to Adam J. Kleinschmit, akleinschmit@dbq.edu.

The authors declare no conflict of interest.

Received 24 August 2023
Accepted 14 November 2023
Published 12 January 2024

testing of how CURE-specific components (e.g., discovery, iteration) relate to anticipated student outcomes more generally (11–15). Understanding the relationship between anticipated long-term student outcomes and CURE elements specific to classroom research will drive CURE development to focus on those elements with the greatest long-term impact (6, 10, 12, 16–18). There remains a need, however, for CURE implementers to assess short-term outcomes as affirmation for initiating or continuing use of a CURE. Evidence that students learn key concepts or science process skills associated with a specific CURE can impact the decision to implement or provide evidence for departmental adoption (19). A number of previously developed instruments may be appropriate for assessing general affective learning domain outcomes associated with research such as critical thinking, ownership, and understanding the process of science [see reference (19)] but are likely too general for assessment of CURE-specific concepts and science process skills.

Nationally disseminated or "network" CUREs are those designed and packaged for dissemination to other instructors, usually at a variety of institutions (6, 20, 21). Assessment of concepts or science process skills associated with a specific network CURE can be challenging because an instrument developed by one instructional team for a particular student population may not be broadly applicable across all participating institutions or classrooms. This issue is confounded further by widely adaptable modular CUREs in which the course context may vary as well. Modular network CUREs provide a suite of independent classroom modules that allow instructors to provide students with a taste of authentic research through a brief CURE over two to three laboratory sessions or through full semester-long projects [e.g., references (21–27)]. Short-duration CUREs with one or more modules may lower implementation barriers (28–31) and provide some of the gains associated with long-duration CUREs (13, 27, 32–34).

Some multi-institution CUREs have published results of perceived (e.g., self-reported) or objective student outcomes at a single institution [e.g., references (34–38)], while other network CUREs have reported student perception of outcomes across multiple institutions [e.g., references (13, 39–48)]. In a literature search, we were able to identify only five multi-institution CUREs that have measured objective outcomes of student learning across multiple institutions with a focus on semester-long implementation. These groups published student knowledge outcomes using surveys designed by participating faculty, but none provided specific details describing survey design, development, or validation (42, 44, 45, 49–53). Objective metrics other than learning have been reported across different implementation contexts at the same institution. For example, participation in the Freshman Research Initiative has been associated with increased student retention (54, 55).

The Prevalence of Antibiotic Resistance in the Environment (PARE) project is a library of modular CURE curricula organized around the theme of environmental surveillance of antibiotic resistance (24, 25, 32). The majority of instructors begin by implementing the "core" module but then go on to expand the duration of research in subsequent implementations through addition of other PARE modules (Genné-Bacon, submitted). A goal is to increase student understanding of how presence of antibiotics in the environment can lead to natural selection for resistant microbes. In the core module, designed around eight content and science process skills outcomes (see Results), students form a hypothesis about which local areas might harbor high resistance levels, and then they use standardized methods to collect soil, capture GPS location, perform serial dilutions, and plate onto media with or without a defined concentration of tetracycline. Colonies are counted to determine the number of colony-forming units on each plate type, followed by calculations to determine the percent resistance. The results are uploaded to an international database and results can be visualized in real time on a map (24, 25). The range of institution types participating in PARE, the varying course context, and the differences in total module numbers taught contribute to significant implementation differences across classrooms (25, 32). Because a programmatic goal is to obtain data comparable across a large geographic range during the core module

component, it is important that all students use the same methodology. Therefore, we opted to develop an instrument to exclusively measure learning outcomes associated with the core module, since it is the module common across all or most PARE classrooms and because the methodology and content should be executed nearly the same across classrooms.

The overarching goal of this study was to test a methodological framework for harnessing a CURE community to assist with developing and validating a knowledge and skills assessment instrument. The impetus for this effort stemmed from a majority of PARE instructors indicating that they value knowledge- and skills-based assessments over affective assessments. We polled the PARE instructor network, and 85% of respondents preferred a content- over an affective-based assessment (19). The resulting instrument provides a classroom tool for network instructors and may bolster adoption through access to published efficacy data. The resulting assessment tool and data set have value for those instructors considering adoption in the current education landscape (19), especially at community colleges with rigid curriculum requirements or in settings where administrative support may be a barrier (30, 56, 57). Having a validated CURE-specific instrument fills the void where generalized instruments that measure presumed short-term outcomes, such as developing hypotheses, designing experiments, or project ownership (11, 58–62), may not be suitable for a single module CURE experience. We recognize that focusing on assessing gains in skills or learning does not capture the holistic nature of CUREs as outlined by Brownell and Kloser (16); however, in certain contexts, it may be valuable (19). Thus, we aim to impact a larger scope of undergraduates through authentic classroom research by presenting PARE as a short-duration modular "gateway" CURE accompanied by instructor-valued assessment tools and efficacy data. Here, we outline a CURE network co-design process that can be adopted by other network CUREs to generate a knowledge and skills assessment instrument. We leveraged our faculty network to assess the efficacy of a short-duration CURE module without external funding or a sizable time commitment.

The first objective of the present study was to develop and establish a tool useful for capturing learning gains from a wide diversity of courses and student populations using a CURE network co-design process. The second objective was to measure gains in conceptual knowledge and the ability to analyze and interpret data using this assessment instrument. The second objective was guided by the following research question: Can a network-based short-duration CURE module demonstrate measurable student learning gains in conceptual knowledge and an ability to analyze and interpret data?

METHODS

Faculty co-design collaborators

During the 2017 American Society for Microbiology Conference for Undergraduate Educators meeting, PARE attendees met and decided to establish an assessment group. The entire PARE instructor network was subsequently invited to join the assessment group, which ultimately resulted in the inclusion of seven PARE faculty instructors as part of the co-design process with survey-style feedback from an additional 20 instructors implementing in diverse contexts.

Student participants

Assessment data were obtained from student participants enrolled in courses taught by PARE instructors at varied institutional types and course types from the fall of 2016 through the early spring of 2020 prior to face-to-face courses transitioning to online due to the COVID-19 pandemic. Broad representation of course types and institution types was deliberate to ensure a representative sample of the PARE network.

Data collection procedure

Faculty co-design survey and student assessment data were collected by administering the instrument using the Qualtrics platform. Student participants completed the pre-assessment instrument within a window of up to 7 days prior to implementation of PARE. The post-assessment instrument was completed by student participants within 7 days after completion of the PARE core module. All spring 2020 data were collected prior to COVID-19 social distancing measures taking effect. All protocols used to collect data from human subjects were approved by the Institutional Review Board at Tufts University (IRB# 1511001).

Instrument scoring, item analysis, and iterative design

For multiple-choice items on all versions, item analysis metrics included index of difficulty, item discrimination index, and point-biserial correlation. The index of difficulty was calculated by dividing the number of correct responses by the number of total responses. Item discrimination index was calculated by sorting records based on total assessment score and then subtracting the percent correct from the bottom 27% of all records from the percent correct of the top 27% of all records (63). Point-biserial correlation was calculated by using the "CORREL" array function in Microsoft Excel to determine the correlation coefficient between score on an item and total score on the instrument. Open-ended items on version 1 were scored by two evaluators, and points were assigned for each correct multiple-choice item. Pre-/post-records were combined and assigned a randomized code; two separate individuals scored each item independently using an established rubric. Scores were compared for a subset of items for scoring calibration and to avoid the introduction of scoring artifacts. For scoring of version 2, a portion of responses for each open-ended item was scored using an established rubric. Modifications were made to questions that were not performing well. Specifically, we used a combination of low item discrimination, poor point biserial correlation, and undetectable or negative pre- to post-score differential as criteria for removing or altering questions. In addition, version 1 questions for which a large proportion of instructors and/or students exhibited confusion or misinterpretation based on open-ended text were omitted or revised. In later versions, open-ended questions were converted to a format that could be easily scored (e.g., multiple choice). To convert open-ended questions into multiple-choice versions, we created distractor answers that were commonly observed in the open responses. Overall, a combination of the classroom data analysis, PARE instructor feedback, and instructor perceived difficulty ratings was used to revise the instrument and for establishing instrument validity (64).

Collectively, the evolution of the assessment items can be found in Table S1, and institutional demographics for each pilot can be found in Table S2.

Learning gain data analysis

Learning gains for versions 2–4 are reported for multiple-select, multiple-choice, or calculation items only. Filtering was performed to remove incomplete records and submissions without matching pre-/post-record identifiers. Duplicate entries were also removed. Multiple-choice scoring was automated within Microsoft Excel. Matched pre-/post-item scores were totaled for each participant to compute individual learning gains. During analysis of the composite data set (instrument version 4), we noticed that a notable proportion of pre-/post-records were completed within a suspiciously short time duration. To explore if short-duration records influence detected learning gains, we removed records with a time duration of <400 seconds from pre-/post-assessment records and any associated unmatched records. This resulted in purging 77 records, yielding a total of 571 matched records with >400 seconds in time duration. Additionally, we noticed that a large portion of the version 4 records (~21%) were from multiple sections of a single community college course (Table S2). To investigate if this institution was driving the composite learning gains, we filtered these records and re-analyzed the

data. The resulting filtered data set exhibited a significant *P* value, indicating that this single institution was not driving learning gains (Fig. S1; $P < 0.0001$, two-tailed paired *t*-test, $n = 448$). In addition, a single institution in the version 4 data set was composed of a majority fraction of students that self-identified as "Black or African American." This institution is located in Botswana, where Black Africans likely have a different cultural experience than Black Americans. Considering that we were specifically interested in the experience of historically underrepresented groups in the United States, we ran the general linear mixed model analysis with and without this class for underrepresented minority analysis, but we did not observe statistical significance in either case.

Statistical analysis for data associated with instrument versions 1–4 (paired two-tailed *t*-tests and Cohen's *d* effect size) were completed using JASP (version 0.14.1; JASP, 2020(65)) . To compare test scores between various demographic groups, we used a generalized linear mixed model with institution as the random factor, post-test score as the dependent variable, demographic or course metadata category (e.g., race, institution type) as the fixed factor, and pre-test score as the covariate. This analysis was performed on instrument version 4 using Statistical Package for the Social Sciences (SPSS) (IBM SPSS version 7 for Mac). This analysis detects differences in post-test score between groups while controlling for pre-test score and institutional random effects (to account for the nested structure of the data across multiple institutions). Partial η^2 is used as a measure of effect size. Because participating institutions do not vary in classification, no random effects nesting was used for the analysis of institution-type differences in test scores.

RESULTS

Network co-design of an assessment instrument for diverse classroom contexts

The main objective of the assessment instrument development process was to produce an assessment that could be used across a wide diversity of courses and student populations within the PARE network. To attain this goal, we conceived a co-design methodological framework for instrument development and validation (Fig. 1). The final product of this structured validation process is an instrument that can be efficiently scored for scaled data collection in courses implementing PARE in diverse contexts. The process entailed generation of four iterations of the instrument and data collection

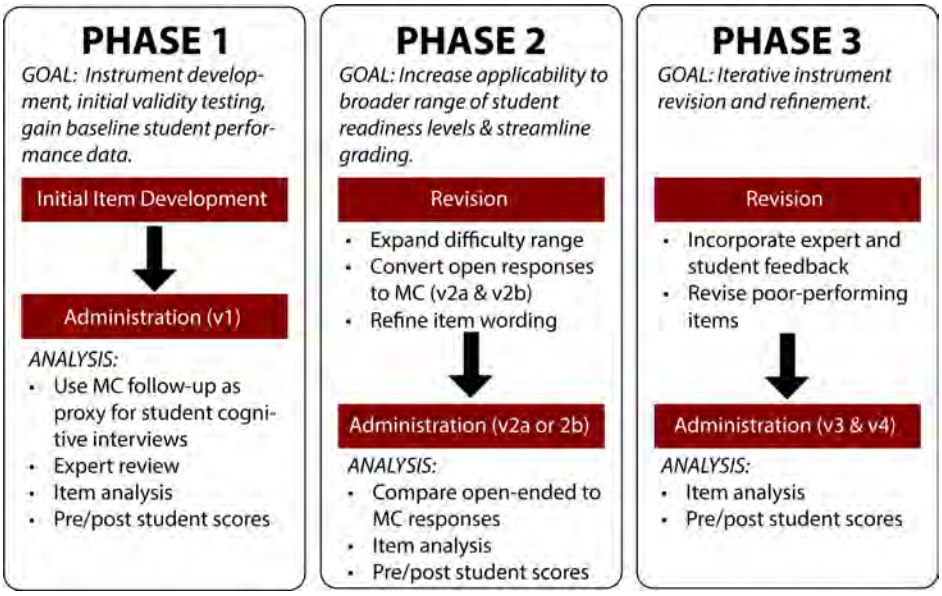

FIG 1 Workflow of the co-design methodological framework for assessment instrument development and validation (see text for details). MC = multiple choice

TABLE 1 Summarized assessment data for instrument versions 1–4[a]

Instrument version	Number of institutions	Matched pre-/post-assessment records	Institution type				
			High school	Associate's colleges	Baccalaureate colleges	Master's colleges	Doctoral universities
1	4	45	0	5	12	13	15
2A	11	253	0	100	78	28	47
2B	10	276	14	41	99	65	57
3	7	254	0	214	36	4	0
4	21	571	0	206	359	3	80
Total	53	1,370	14	566	584	113	180

[a]Versions 2A and 2B were used in different sections of the same course for some participating institutions. In a few circumstances, a single instructor has implemented PARE in two separate courses at the same institution.

TABLE 2 Courses participating in assessment data collection for versions 1–4[a]

Course focus	Course student composition				Course level	
	STEM majors	Mixed majors	Allied health	Non-majors	Upper	Lower
Microbiology	17	9	10	NA	10	26
Genetics	2	NA	NA	NA	NA	2
Biology	8	2	1	3	NA	14
Total	27	11	11	3	10	42

[a]STEM majors includes courses composed of strictly biology majors and courses that enroll a wide diversity of STEM majors. NA, not applicable.

over the period of three academic years. Data collection efforts have involved over two dozen unique institutions (Table 1), 42 course offerings (Table 2), and over 1,300 pre-/post-matched assessment record data points (Table S2). Due to the adaptability of the PARE Project as a short-duration CURE, participating classrooms included diverse environments such as non-biology majors, pre-allied health majors, comprehensive STEM majors, mixed majors, and biology majors. Items were targeted to reach across an item difficulty range to accommodate lower-level and upper-level PARE-implementing courses within the network (Table 2).

Phase 1: instrument development and initial validity testing

PARE learning objectives (Table 3) were used to develop a set of 14 questions (24) consisting of a combination of open-ended and multiple-choice style questions. This initial instrument was used to collect pilot data from four institutions and course offerings. Statistically significant learning gains from this version were previously reported ($P < 0.0001$; two-tailed paired *t*-test; $n = 45$) (24). To assess whether students were interpreting the questions as intended, we provided a field for students to explain their answer choice after each multiple-choice question. This was used as a metric to assess face validity and to identify poorly worded questions that were then flagged for revision. Commonly chosen wrong answers were flagged as potential misconceptions and were used in subsequent iterations as distractors.

Variation in performance based on institution prompted us to seek input from faculty who teach the PARE Project. Version 1 of the instrument was converted to a Qualtrics-based survey and distributed to PARE faculty to assess content validity as well as perceived difficulty, appropriateness for their student population, and general clarity. A total of 27 instructors (Fig. 2A) took the survey and offered feedback. The items on this version were identical to the student version but included an open-response space after each question for the instructor experts to offer general feedback and a Likert scale question to capture perceived difficulty level for their particular student audience.

Instructor survey respondents expressed a range of opinions regarding perceived difficulty level and opinions differed across the different items (Fig. 2B). For example, a single item requiring math analogous to what is required in the core module was perceived as being overly simplistic and not appropriate for collegiate level to being too difficult for students without context on antimicrobial resistance. This was our

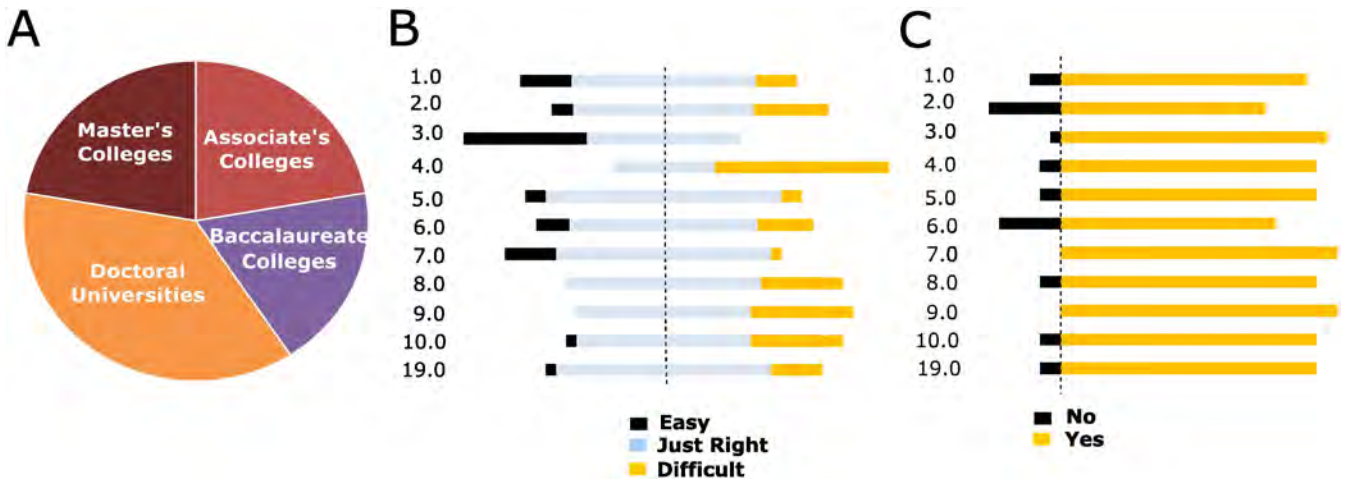

FIG 2 Faculty feedback survey data for version 1 of the assessment instrument. (A) Institution-type representation of faculty responses (n = 27). (B) Faculty responses regarding perceived item difficulty for their specific student population. (C) Faculty responses to survey question "would this assessment item be appropriate for your course content?".

TABLE 3 PARE learning objectives for the core PARE module mapped to general student learning outcomes

Core PARE Project student learning objectives (students should be able to…)	General student outcomes from CURE participation	Alignment to vision and change core competencies[a]
1) Express and convert numerical values between fractional, decimal, and scientific notation.	Increased content knowledge	LO 1–2: quantitative reasoning—numeracy
2) Calculate the number of colony-forming units per gram of soil.		
3) Explain the rationale and process for performing serial dilutions on microbiological samples.		LO 3: process of science—study design
4) Explain how antibiotics can provide a selective pressure influencing natural selection of microbial populations.		LO 4–5: interdisciplinary nature of science—connecting scientific knowledge
5) Describe the potential implications for human health posed by the presence of antibiotics in the environment.		LO 5: science and society—science's impact on society
6) Represent or interpret a given set of authentic ("noisy") data in a table, graph, etc.	Ability to analyze and interpret data	LO 6: quantitative reasoning—quantitative and computational data analysis
7) Reflect on unexpected experimental results and determine nature of error/troubleshoot.		LO 7: process of science—study design
8) Describe the importance of experimental replication and the ability to be cautiously skeptical of data.		LO 8: process of science—data interpretation and evaluation
9) Demonstrate evidence of the ability to recognize how "scientists think" by conceptualizing data that contradict experimental predictions in response to a situational prompt question concerning unexpected experimental results.	Conceptions of what it means to think like a scientist	LO 9: (i) Modeling and simulation—model application (ii) Process of science—scientific thinking

[a]Vision and change core competencies as described by Clemmons et al. (2020) (66).

goal, considering the diverse target student population being assessed. Item analysis of instrument version 1 (Table S1), indicated that the index of difficulty mirrored the general trend of difficulty ratings provided by instructors. Overall, the participating instructor respondents holistically rated version 1 instrument items as being appropriate for their course context by a large majority (Fig. 2C).

Phase 2: expansion of difficulty range and conversion of open-ended to multiple-choice items

The goals of phase 2 were to (i) add additional questions to generate both a relatively novice and challenging question for each learning objective and (ii) convert

open-ended questions to an easily scorable format for streamlined analysis. Instructor survey feedback on version 1, coupled with student pilot data, guided instrument refinement. Item analysis was used to identify learning objectives for which additional item(s) or modification was necessary to provide varying levels of difficulty and/or to improve clarity. To scale up administration to large cohorts, we converted open-ended items to multiple choice by using misconceptions articulated in open-ended responses to generate distractors for multiple-choice items. For example, an open-ended item and its multiple-select counterpart are shown below (see box) followed by some example incorrect student responses that informed development of the distractor choice "c" in the multiple-select version.

Open ended

A biologist wants to investigate the bacteria living in a sample of pond water. She performs a serial dilution of the pond water then plates each dilution on appropriate growth medium. What is the purpose of performing serial dilutions on the pond samples? Why can't she just plate the pond sample directly on the medium?

Multiple select

A biologist wants to investigate the bacteria living in a sample of pond water. She performs a serial dilution of the pond water then plates each dilution on appropriate growth medium. What is the purpose of performing serial dilutions on the pond samples? Why can't she just plate the pond sample directly on the medium? Select ALL that apply.

a. The bacteria in the original sample are likely too concentrated, thus dilutions will reduce the number of cells per milliliter for counting distinguishable colonies.
b. Diluting lets you make an accurate plate count by creating space between viable cells.
c. This step removes contaminants and waste products from the lake water that inhibit bacterial growth.
d. Diluting helps concentrate the bacterial cells in samples to provide more accurate plate counts.

Examples of incorrect responses that led to generation of incorrect multiple select distractor choice "c"

- By performing the serial dilutions it separates the bacteria from the sample of pond water which will make the bacteria grow more effectively.
- Honestly I don't know, probably to remove unneeded chemicals and waste products.
- She wants to isolate the bacteria as much as possible in just pure water, but plating the water straight from the lake would mean that contaminates from the lake itself would also get plated.
- To search for contamination before she places it on the medium.

We compared performance on version 1 to the modified counterpart item on versions 2A and 2B. Multiple-choice, multiple-select, and calculation items that exhibited a combination of low item discrimination, poor point-biserial correlation, and undetectable or negative post-/pre-score differential metrics were flagged for re-wording or removal from the instrument. Overall, we did not see a major difference in performance on open-ended items that were converted to one of the easily scorable formats.

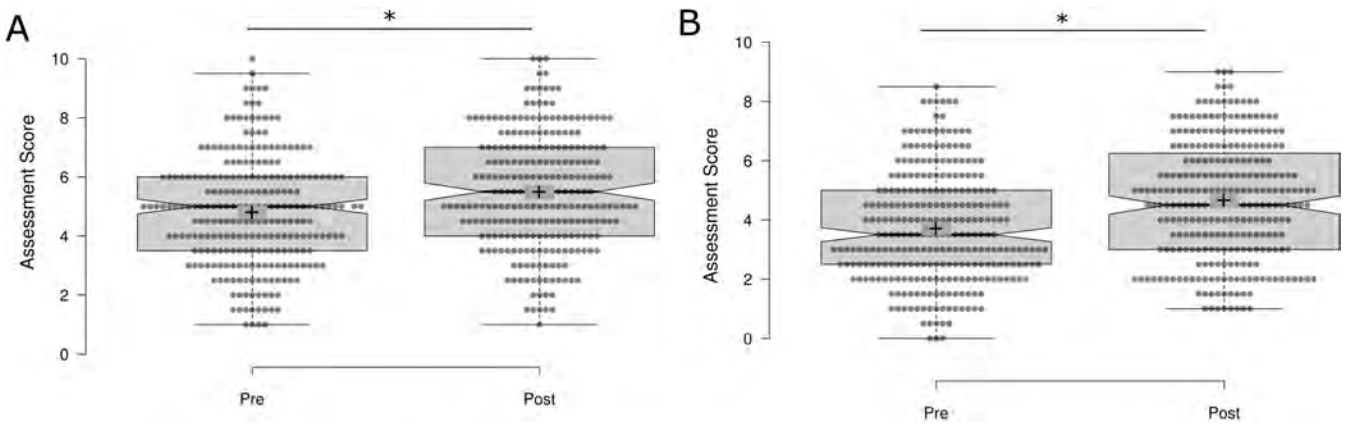

FIG 3 Aggregate matched pre-/post-assessment PARE quiz scores (version 2) indicate significant participant learning gains on multiple-choice questions. Two versions of the MC/MS portion (10 items) of the assessment were administered pre- and post-implementation of the PARE curriculum during the fall 2018 semester. (A) Nine cohorts of student participants (n = 253) at independent institutions. A two-tailed paired t-test indicated significant difference (P < 0.0001) between pre- (4.80) and post-quiz (5.49) means. (B) Nine cohorts of student participants (n = 276) at independent institutions. A two-tailed paired t-test indicated significant difference (P < 0.0001) between pre- (3.70) and post-quiz (4.65) means. + = mean; box notches = ±1.58*IQR/sqrt(n) and represent the 95% confidence interval for each median. Non-overlapping notches give roughly 95% confidence that two medians differ. * = statistical significance (P < 0.0001).

The resulting data set of responses from version 2 represented 15 unique institutions implementing PARE in diverse institution types, courses, student academic interests, and course levels and yielded 529 matched pre-/post-assessment records (Tables S1 to S3).

Analysis of the scored multiple-choice, multiple-select, and calculation items (10 total items) and the resulting total instrument mean for matched pre-/post records indicated significant learning gains for both forms "A" (n = 253) and "B" (n = 276) (Fig. 3; P < 0.0001; two-tailed paired t-test).

Phase 3: iterative instrument revision and refinement

The data set from version 2 pilots was skewed toward baccalaureate institutions with an undergraduate focus, so we embarked on an additional round of pilots with a revised version 3 assessment instrument, consisting of only multiple-choice, multiple-select, or calculation questions (15 items total) (Table S1 to S3).

Having multiple questions per learning outcome assisted with item optimization. For example, multiple-choice question (item 12.0) in version 2 form "A" lacked a mean difference between pre/post (Table S1); thus, it was dropped in favor of the alternative, better-performing item 13.0 addressing the same learning objective. Despite attempts to systematically improve items, this proved difficult in some cases. For example, efforts to improve student performance on item 5.2 were unsuccessful. Interestingly, some instructors indicated in the survey that their cohorts of students with remedial math skills would likely be more challenged with the question. This question tested basic understanding of scientific notation. It was still retained in the instrument, as it was deemed important to keep a question addressing this established learning outcome.

Versions 3 and 4 were both composed entirely of easily scorable items (multiple-choice, multiple-select, and two calculation items) for scalability and balanced brevity with coverage of the aligned student learning outcomes. Item analysis metrics indicate that items in the final instrument (version 4) span a range of difficulty while providing good discrimination ability (Table 4). The pre/post mean differential was indicative of significant learning gains (Fig. 4; P < 0.02; two-tailed paired t-test).

TABLE 4 Assessment instrument version 4 item analysis data

LO	Item ID	Format[a]	Version	Status relative to previous version	Item difficulty	Item discrimination	Point-biserial correlation	Mean post – mean pre
	1.1	C	4	Converted to calc.	0.61	0.58	0.49	−0.03
1	2.2	C	4	Converted back to calc.	0.52	0.69	0.58	0.03
	3.2	MC	4	Same as v3	0.44	0.33	0.29	0.06
2	4.0	MC	4	Same as v1	0.31	0.42	0.38	0.08
3	5.2	MC	4	Same as v3	0.44	0.46	0.35	0.04
4	8.1	MC	4	Same as v2/3	0.56	0.57	0.45	−0.01
5	13.0	MC	4	Same as v2b/3	0.37	0.51	0.43	0.05
	14.1	MC	4	Same as v3	0.23	0.27	0.27	0.01
6	18.0	MC	4	Same as v2b	0.3	0.35	0.32	0.04
	19.0	MC	4	Same as v1	0.46	0.56	0.47	0.06
7	21.1	MC	4	Same as v3	0.38	0.58	0.48	0.1
8	22.1	MC	4	Same as v3	0.79	0.49	0.47	0.01

[a]MC = multiple choice; C = calculation; OE = open ended.

Assessment of a short-duration CURE

Participating in the PARE core research module leads to learning gains

To determine if a network-based short-duration CURE module can demonstrate measurable student learning gains, data collection for version 4 of the assessment instrument was ramped up to collect data from a wider array of institutions, course types, and course levels, which included 21 unique institutions and yielded a total of 571 matched pre-/post-records (Table 4; Table S1–S3). The aggregate data set from version 4 of our assessment instrument showed significant learning gains (Fig. 5; $P < 0.0001$; two-tailed paired *t*-test). To measure the magnitude of this shift in pre-/post-assessment score we calculated Cohen's *d*, which indicated a small effect size ($d = 0.243$). This result

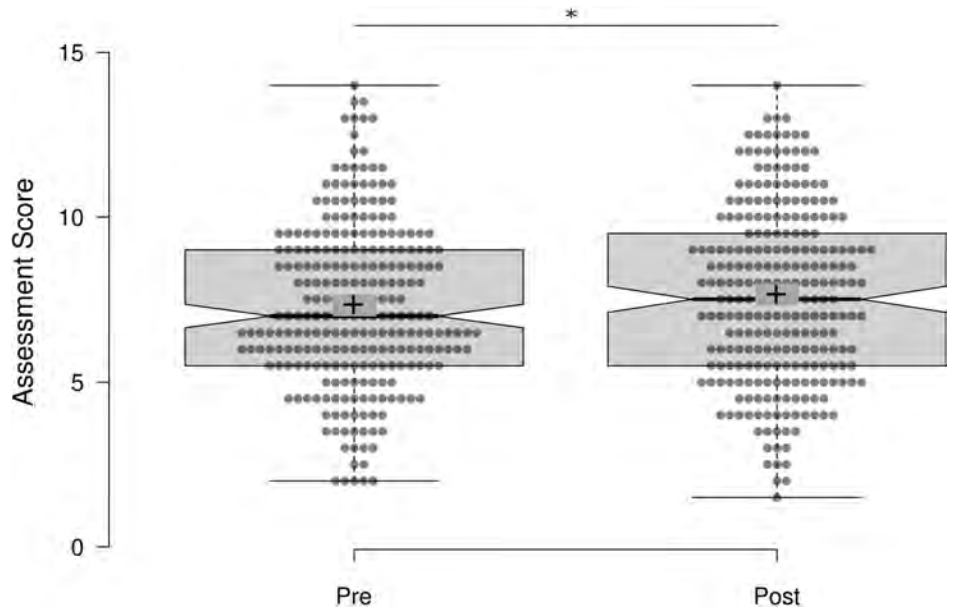

FIG 4 Matched pre-/post-assessment PARE quiz scores (Version 3) indicate significant participant learning gains on MC/MS/C questions. Seven cohorts of student participants ($n = 254$) at independent institutions completed the assessment instrument. A two-tailed paired *t*-test indicated significant difference ($P < 0.02$) between pre- (7.33) and post-quiz (7.64) means. + = mean; box notches = ±1.58*IQR/sqrt(n) and represent the 95% confidence interval for each median. Non-overlapping notches give roughly 95% confidence that two medians differ. The box plot visual was generated using BoxPlotR (Spitzer et al., 2014) (67). * = statistical significance ($P = 0.02$).

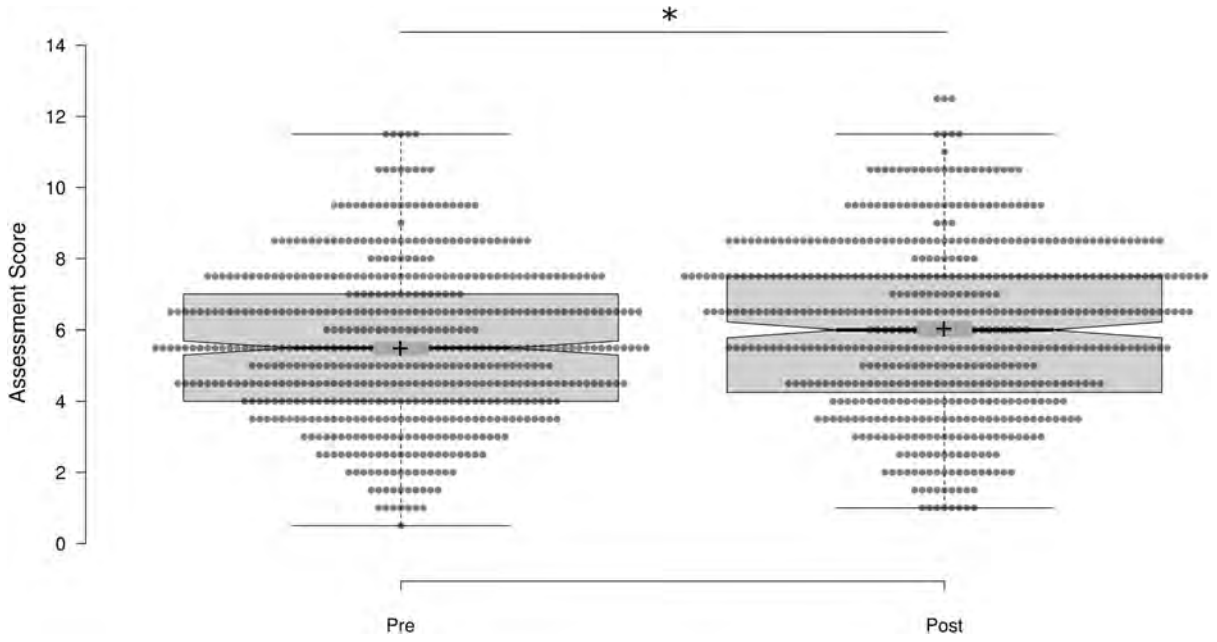

FIG 5 Matched aggregated data (version 4) indicate significant learning gains. The black cross bar indicates the aggregate mean. Filled gray boxes indicate 95% confidence intervals of the mean. Mean score (5.48 pre; 6.02 post), $r = 0.5234$, Cohen's $d = 0.243$, $n = 571$. * = statistical significance ($P < 0.0001$).

is not surprising considering that the core PARE research module is implemented in two laboratory sessions within a range of course types with diverse classroom learning objectives.

Mode of administration's influence on the data set

The pre- and post-assessment instruments were administered consistently within 7 days before and after completing the CURE module, respectively, but the mode of administration varied as it was left to the instructor's discretion. Some instructors chose to provide structured class time, while others directed student participants to complete the survey on their own time. We were interested if the mode of administration influenced pre-/post-assessment score differentials. Learning gains were significantly higher for assessments completed within a structured environment compared to outside of class (Fig. S2, p; $P = 0.008$, two-tailed *t*-test). Additionally, we noted that of the records that were previously filtered for taking <400 seconds to complete, >76% of these filtered records were from participants that took the assessment outside of the classroom environment. Asynchronously administered records included the same associate degree-granting college that dominated the version 4 data set.

Participation in the PARE core research module leads to conceptual and competency-based learning gains

We investigated whether students exhibited learning gains in content knowledge (items associated with LOs 1–5) independently of the ability to analyze and interpret data (items associated with LOs 6–8). Mean aggregate pre-/post-assessment scores for each category exhibited significant learning gains (Fig. 6; $P < 0.0001$, small effect size, $d = 0.185$, and $P < 0.0001$ respectively; two-tailed paired *t*-test; $n = 571$; small effect size, $d = 0.216$).

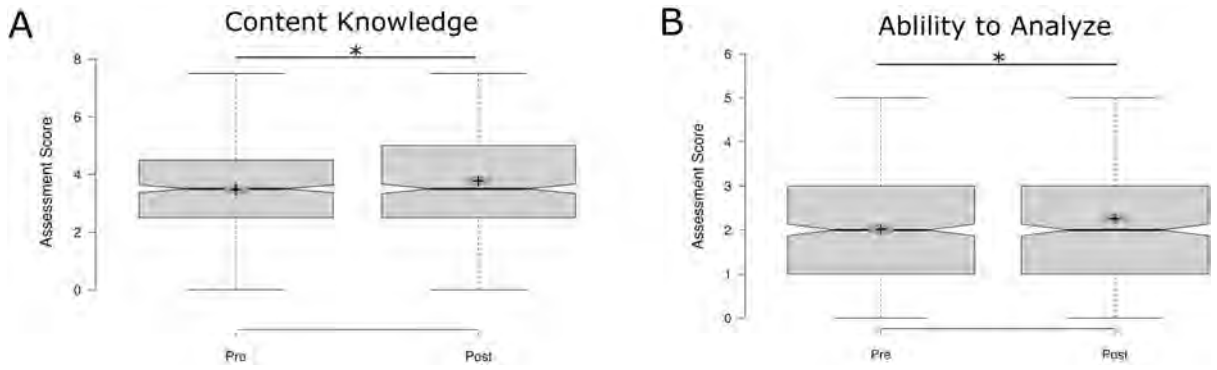

FIG 6 Significant learning gains are observed with both (A) increased content knowledge and (B) ability to analyze and interpret data-categorized assessment questions (filtered for <3,000-second records). Mean score of content knowledge items (3.469 pre; 3.770 post); ability to analyze items (2.012 pre; 2.252 post). * = statistical significance ($P < 0.0001$).

TABLE 5 Generalized linear mixed model statistical analysis summary ($n = 568$)[a]

Demographic variable[b]	*P*-value	Bonferroni α[c]	Effect size (η^2)[d]
Institutional classification	0.007	0.0125	0.018 (small)
Gender identity	0.012	0.0125	0.009
Underrepresented minority	0.911	0.0125	0
First-generation college student	0.891	0.0125	0.001

[a]Dependent variable: post-test score; co-variant: pre-test score; random effects: institution name; fixed effect: variable being tested.
[b]Institutional classification: deleted three records from master's colleges and compares associate's colleges, baccalaureate colleges, and doctoral universities; gender identity: removed all entries with no gender or non-binary; underrepresented minority: excludes Asian; first-generation college student: compares the following three categories: both parents attended college, one parent attended college, and neither parent attended college.
[c]Adjusted α (α = 0.05) obtained by dividing α by the four generalized linear mixed model tests performed.
[d]Partial η^2 metric was used for effect size.

Learning gains are independent of demographic variables

Next, we sought to investigate if student learning gains were independent of contextual factors associated with implementation of PARE across diverse classrooms. Learning gains were independent of underrepresented minority status (Table 5; generalized linear mixed model, $P = 0.911$, $\eta^2 = 0$) and first-generation college student status (Table 5; $P = 0.891$, $\eta^2 = 0.001$). We did detect that learning gains were weakly associated with institution classification (Table 5; $P = 0.007$, $\eta^2 = 0.018$) and gender identity (Table 5; generalized linear mixed model, $P = 0.012$, $\eta^2 = 0.009$), although institutional classification and gender identity both exhibited relatively small effect size.

DISCUSSION

Leveraging a network for iterative instrument co-design

Network CUREs have been documented for providing a centralized support system to bring together diverse intellectual capital across institutional types to drive scientific and curriculum structure [e.g., references (21, 22, 45, 46, 50, 52, 68–73)]. Among network CUREs, there are examples of specific institutional faculty members developing "in house" objective learning gain assessments for use within their own or a few institutions (37, 38, 51, 74). Some network CUREs have also used published instruments for objective measurement of critical thinking (38) and psychological outcomes associated with persistence (40). Furthermore, a few CURE networks have reported objective content learning gains across many institutions, although details on the instrument design were not reported (29, 44, 50, 52). Engagement of the wider network for collaborative design and validation of an assessment instrument has been underutilized. This unexplored

avenue of leveraging the collective CURE network for development and validation of an assessment instrument could be especially fruitful for networks with limited grant support and resources traditionally used to drive these important efforts. The iterative co-design process established here democratizes assessment item development and modification, with feedback from faculty representing all stakeholders within the broad network. This process provides the means for validation through collection of many data points iteratively over time with diverse faculty experts helping craft the assessment instrument without the need for external grant support. We are not aware of any other published, validated instruments to measure learning gains in a network CURE. The development of this instrument not only fills a need within the PARE network, but the documented process may also guide other CURE networks to leverage their support systems in developing validated knowledge- and skills-based assessments that meet their specific learning and program goals.

Challenges of developing an assessment instrument for a broadly implemented short-duration CURE

It is impractical for faculty running a CURE developed for sole use within their own class to validate a learning gains assessment, but for network CUREs, evidence of positive outcomes is valuable for other faculty considering adoption or to gain administrative support (19). In many cases, having limited resources to put toward developing the instrument can also create a barrier. Since student participants in lower-division CUREs are heterogeneous, especially in mixed majors courses, it can also be challenging to develop an assessment instrument for use across varied implementation contexts. We encountered three major obstacles when considering assessment of PARE learning gains. First, the core PARE module is short duration, implemented across one or two laboratory sessions. This limits the number and depth of assessment items that can be created. To address this challenge, we chose to keep the instrument focused on the core PARE module learning objectives and limited to a few questions, similar to the length of a quiz. Second, our assessment had to measure gains over a diversity of course types (e.g., introductory biology, microbiology, genetics) and for participants with varied vocational interests (e.g., majors, non-majors, allied health majors) and prior life science training (upper division, lower division). For example, in the first version, we observed a learning gain ceiling effect for students with a strong foundation in the concepts and competencies explored. To address this challenge, we gathered input from faculty on their perceived difficulty for the students they teach prior to scaled-up administration and used item analysis subsequently to reduce ceiling effects and to ensure a range of difficulty level of questions. Third, there is variability in implementation. PARE is intentionally flexible and expandable through addition of multiple additional modules, so students at different institutions may experience different durations of the PARE research experience. To mitigate these differences, we limited the assessment instrument items to the stated learning objectives associated with the core module, and we gained feedback from a range of faculty who teach PARE in different contexts to gauge their assessment of concept coverage. We did capture information on the number of modules implemented for each cohort of respondents, but we did not see any significant differences attributable to the quantity of PARE modules completed.

Detection of learning gains in a short-duration CURE

Historically, CURE curricula have been prescribed as semester-long laboratory interventions, and previous studies have pointed toward duration and intensity level of research experiences being paramount for such interventions to be impactful (49, 75). Expectations for student outcomes with such a short-duration CURE such as PARE were minimal, since the focus of PARE is on faculty change, not student outcomes. A short, yet expandable modular CURE is an attractive pedagogical approach for faculty considering the transition from a traditional set of teaching laboratory exercises to authentic

research (31, 76). Use of expandable CURE modules can facilitate moving past implementation barriers by providing the flexibility to progressively expand integration of CURE modules while optimizing module fit with course goals. However, potential implementing faculty still want to know whether their students will learn concepts integral to their course. Aggregate data analyzed independently from the four different iterations of the assessment instrument exhibit marginal but statistically significant learning gains across diverse classroom contexts. Furthermore, our instrument was designed to assess general microbiological principles, not those exclusive to participation in PARE; therefore, we expect students may score higher on our pre-instrument than for a CURE-specific instrument. Given this, it is notable that students did exhibit learning gains. The diverse classroom contexts included implementation in a variety of life science courses (e.g., introductory biology, microbiology, genetics), students with varied vocational interests (e.g., STEM majors, allied health, non-majors), level of training within the biological sciences (lower-division, upper-division courses), and institution type (Table 2). Our results provide evidence for faculty interested in transitioning to classroom research that students will still learn skills associated with a traditional microbiology laboratory course.

Study limitations

It is worth noting that we observed some variance within the item analysis metrics associated with non-altered items (e.g., item discrimination) between versions of the instrument. It is plausible that differences in participating student classrooms and institutions for different rounds of data collection may have contributed to these results. For example, pilot data from instrument versions 3 and 4 had a considerable proportion of student participants from a single associate degree-granting college (Table S2). When possible, we sought to test assessment items across multiple versions to obtain multiple data points to assist with instrument validity and refinement.

Value of learning gain assessment

Faculty may require or desire evidence of particular learning or skills outcomes prior to adoption of a new CURE. For example, PARE network faculty value evidence of knowledge- and skills-based outcomes over attitudinal evidence. The reasons for this are varied, but a prominent reason is that they are useful for demonstrating that a CURE meets course requirements including community college transfer agreements (19). In addition, learning gain evidence from this study adds to current CURE logic models (10, 12) by providing evidence of the utility of a short-term CURE. Specifically, this study suggests that increased content knowledge and improved analytical skills may be short-term outcomes associated with CUREs. Current outcome models focus on semester-long CUREs in advanced majors courses (10, 12); here, we demonstrate that a short-duration CURE also supports this portion of the CURE logic model, independent of student demographics. PARE represents a short-duration low-cost modular CURE, which is easily adaptable at all undergraduate divisional levels, while yielding effective knowledge gains. Effectiveness of CUREs, especially at institutions with limited resources such as community colleges, may depend on availability of an easily adoptable curriculum, use of common tools, and ability for expansion over time. In other words, PARE represents what can be a long-term pedagogical CURE project that does not require constant curricular change by instructors but still produces novel data and has value to the greater scientific community.

ACKNOWLEDGMENTS

This manuscript is based upon work supported by the National Science Foundation Grant #1640399 awarded to C.B.-S.

We would like to thank the following PARE faculty for providing survey-style feedback on the instrument and/or facilitating assessment data collection: Jessica Allen, Simran

Banga, Tim Beagley, Peggy Brickman, Stacie Brown, John Buchner, Mike Buckholt, Jeremy Buegge, Fang Cindy Xie, Megan Cole, Lubomira Cubonova, Siaumin Fung, Brittany Gasper, Brad Goodner, Sharon Gusky, Chantal Koechli, Renu Kumar, Archana Lal, Matthew Mastropaolo, Veronica Neumann, Jim Nienow, Selene Nikaido, Roxanna Ochoa, Sarah Olken, Sumali Pandey, Alisa Petree, Rachel Pritchard, Erin Richard, Christine Sestero, Sarah St. Onge, Nikolas Stasulli, Patricia Storlie, Zoe Thompson, Sandra Thompson-Jaeger, and Eric Warrick.

AUTHOR AFFILIATIONS

[1]Department of Natural and Applied Sciences, University of Dubuque, Dubuque, Iowa, USA
[2]Department of Medical Education, Tufts University School of Medicine, Boston, Massachusetts, USA
[3]Department of Biology, Birmingham-Southern College, Birmingham, Alabama, USA
[4]Department of Biology, San Francisco State University, San Francisco, California, USA
[5]Department of Biological and Environmental Sciences, Capital University, Columbus, Ohio, USA
[6]Department of Biology, University of Wisconsin-River Falls, River Falls, Wisconsin, USA

AUTHOR ORCIDs

Adam J. Kleinschmit http://orcid.org/0000-0003-3820-524X
Jennifer R. Larson http://orcid.org/0000-0003-4352-3517
Carol Bascom-Slack http://orcid.org/0000-0002-2950-8997

AUTHOR CONTRIBUTIONS

Adam J. Kleinschmit, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing | Elizabeth Genné-Bacon, Conceptualization, Formal analysis, Methodology, Writing – original draft, Writing – review and editing | Kevin Drace, Investigation, Methodology, Writing – review and editing | Brinda Govindan, Investigation, Methodology, Writing – review and editing | Jennifer R. Larson, Investigation, Methodology, Writing – review and editing | Amber A. Qureshi, Investigation, Methodology, Writing – review and editing | Carol Bascom-Slack, Conceptualization, Formal analysis, Funding acquisition, Methodology, Visualization, Writing – original draft, Writing – review and editing

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental material (jmbe00149-23-S0001.docx). Tables S1 to S3, Fig. S1 and S2, and assessment instruments.

REFERENCES

1. Kuh G. 2008. High-impact educational practices: what they are, who has access to them, and why they matter. AAC&U. https://www.aacu.org/publication/high-impact-educational-practices-what-they-are-who-has-access-to-them-and-why-they-matter.
2. Intemann K. 2009. Why diversity matters: understanding and applying the diversity component of the national science foundation's broader impacts criterion. Soc. Epistemol. 23:249–266. https://doi.org/10.1080/02691720903364134
3. Handelsman J, Elgin S, Estrada M, Hays S, Johnson T, Miller S, Mingo V, Shaffer C, Williams J. 2022. Achieving STEM diversity: fix the classrooms. Science 376:1057–1059. https://doi.org/10.1126/science.abn9515
4. 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
5. National Academies of Sciences, Engineering, and Medicine (NASEM). 2015. Integrating Discovery-Based Research into the Undergraduate Curriculum: Report of a Convocation. National Academies Press,Washington, D.C. http://www.nap.edu/catalog/21851.

6. Dolan. 2016. Course-based undergraduate research experiences: current knowledge and future directions. Nat Res Council 1:1–34. https://doi.org/10.1187/cbe.14-10-0167
7. Dolan WG. 2021. A guide to course-based undergraduate research. Macmillan Higher Education.
8. Hensel NH, Davidson CN. 2023. Course-based undergraduate research. In Course-based undergraduate research: educational equity and high-impact practice. Taylor & Francis, New York.
9. Buchanan AJ, Fisher GR, Dolan EL. 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
10. 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
11. Hanauer DI, Dolan EL. 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
12. Corwin LA, Graham MJ, Dolan EL. 2015. Modeling course-based undergraduate research experiences: an agenda for future research and evaluation. CBE Life Sci Educ 14:es1. https://doi.org/10.1187/cbe.14-10-0167
13. Mader CM, Beck CW, Grillo WH, Hollowell GP, Hennington BS, Staub NL, Delesalle VA, Lello D, Merritt RB, Griffin GD, Bradford C, Mao J, Blumer LS, White SL. 2017. Multi-institutional, multidisciplinary study of the impact of course-based research experiences. J Microbiol Biol Educ 18:18–2. https://doi.org/10.1128/jmbe.v18i2.1317
14. Corwin LA, Runyon CR, Ghanem E, Sandy M, Clark G, Palmer GC, Reichler S, Rodenbusch SE, Dolan EL. 2018. Effects of discovery, iteration, and collaboration in laboratory courses on undergraduates. CBE Life Sci Educ 17:ar20. https://doi.org/10.1187/cbe.17-07-0141
15. 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
16. Brownell SE, Kloser MJ. 2015. Toward a conceptual framework for measuring the effectiveness of course-based undergraduate research experiences in undergraduate biology. Stud. High. Educ. 40:525–544. https://doi.org/10.1080/03075079.2015.1004234
17. Linn MC, Palmer E, Baranger A, Gerard E, Stone E. 2015. Education undergraduate research experiences: impacts and opportunities. Science 347:1261757. https://doi.org/10.1126/science.1261757
18. Beardslee J. 2021. The effects of course-based undergraduate research experiences on student retention, graduation rates, and research interest in STEM PhD Thesis, Southern Illinois University at Edwardsville
19. Kleinschmit AJ, Govindan B, Larson JR, Qureshi AA, Bascom-Slack C. 2023. The CURE assessment landscape from the instructor's point of view: Knowledge and skills assessments are highly valued support tools for CURE adoption. Front Educ 8:1291071. https://doi.org/10.3389/feduc.2023.1291071
20. Shortlidge EE, Brownell SE. 2016. How to assess your CURE: a practical guide for instructors of course-based undergraduate research experiences. J Microbiol Biol Educ 17:399–408. https://doi.org/10.1128/jmbe.v17i3.1103
21. Connors PK, Lanier HC, Erb LP, Varner J, Dizney L, Flaherty EA, Duggan JM, Yahnke CJ, Hanson JD. 2021. Connected while distant: networking cures across classrooms to create community and empower students. Integr Comp Biol 61:934–943. https://doi.org/10.1093/icb/icab146
22. Elgin SCR, Hauser C, Holzen TM, Jones C, Kleinschmit A, Leatherman J, Genomics Education Partnership. 2017. The GEP: crowd-sourcing big data analysis with undergraduates. Trends Genet 33:81–85. https://doi.org/10.1016/j.tig.2016.11.004
23. Dizney L, Connors PK, Varner J, Duggan JM, Lanier HC, Erb LP, Flaherty EA, Yahnke CJ, Hanson JD, Connors PK, Varner J, Duggan JM, Lanier HC, Erb LP, Flaherty EA, Yahnke CJ, Hanson JD, Connors PK, Varner J, Duggan JM, Lanier HC, Erb LP, Flaherty EA, Yahnke CJ, Hanson JD. 2021. An introduction to the squirrel-net teaching modules. CourseSource 7. https://doi.org/10.24918/cs.2020.26
24. Genné-Bacon EA, Bascom-Slack CA. 2018. The PARE project: a short course-based research project for national surveillance of antibiotic-resistant microbes in environmental samples. J Microbiol Biol Educ 19:10. https://doi.org/10.1128/jmbe.v19i3.1603
25. Fuhrmeister ER, Larson JR, Kleinschmit AJ, Kirby JE, Pickering AJ, Bascom-Slack CA. 2021. Combating antimicrobial resistance through student-driven research and environmental surveillance. Front Microbiol 12:577821. https://doi.org/10.3389/fmicb.2021.577821
26. Reyna N, Hensley LL. 2023. CBEC - cell biology education consortium (RCN-UBE introduction). RCN-UBE community. QUBES. https://doi.org/10.25334/7ZJZ-W895
27. Hanauer DI, Nicholes J, Liao FY, Beasley A, Henter H, Gardner S. 2018. Short-term research experience (SRE) in the traditional lab: qualitative and quantitative data on outcomes. CBE Life Sci Educ 17:ar64. https://doi.org/10.1187/cbe.18-03-0046
28. Staub N, Blumer LS, Beck CW, Delesalle VA, Griffin, GD, Merritt RB, Hennington BS, Grillo WH, Hollowell GP, White SL. 2016. Course-based science research promotes learning in diverse students at diverse institutions. CUR Quarterly 38. https://doi.org/10.18833/curq/37/2/11
29. Lopatto D, Hauser C, Jones CJ, Paetkau D, Chandrasekaran V, Dunbar D, MacKinnon C, Stamm J, Alvarez C, Barnard D, et al. 2014. A central support system can facilitate implementation and Sustainability of a classroom-based undergraduate research experience (CURE) in Genomics. CBE Life Sci Educ 13:711–723. https://doi.org/10.1187/cbe.13-10-0200
30. 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. https://doi.org/10.1187/cbe.13-08-0169
31. Genné-Bacon EA, Wilks J, Bascom-Slack C. 2020. Uncovering factors influencing instructors' decision process when considering implementation of a course-based research experience. CBE Life Sci Educ 19:ar13. https://doi.org/10.1187/cbe.19-10-0208
32. Bliss SS, Abraha EA, Fuhrmeister ER, Pickering AJ, Bascom-Slack CA. 2023. Learning and STEM identity gains from an online module on sequencing-based surveillance of antimicrobial resistance in the environment: an analysis of the PARE-seq curriculum. PLoS One 18:e0282412. https://doi.org/10.1371/journal.pone.0282412
33. Dahlberg C, Lee S, Leaf D, Lily L, Wiggins B, Jordt H, Johnson T. 2019. A short, course-based research module provides metacognitive benefits in the form of more sophisticated problem solving. J Coll Sci Teach 48:22–30. https://doi.org/10.2505/4/jcst19_048_04_22
34. Cole MF, Hickman MA, Morran L, Beck CW. 2021. Assessment of course-based research modules based on faculty research in introductory biology. J Microbiol Biol Educ 22:00148–21. https://doi.org/10.1128/jmbe.00148-21
35. Adkins SJ, Rock RK, Morris JJ. 2018. Interdisciplinary STEM education reform: dishing out art in a microbiology laboratory. FEMS Microbiol Lett 365. https://doi.org/10.1093/femsle/fnx245
36. Muth TR, McEntee CM. 2014. Undergraduate urban metagenomics research module. J Microbiol Biol Educ 15:38–40. https://doi.org/10.1128/jmbe.v15i1.645
37. Olimpo JT, Fisher GR, DeChenne-Peters SE. 2016. Development and evaluation of the tigriopus course-based undergraduate research experience: impacts on students' content knowledge, attitudes, and motivation in a majors introductory biology course. CBE Life Sci Educ 15:ar72. https://doi.org/10.1187/cbe.15-11-0228
38. Caruso JP, Israel N, Rowland K, Lovelace MJ, Saunders MJ. 2016. Citizen science: the small world initiative improved lecture grades and California critical thinking skills test scores of nonscience major students at Florida Atlantic University. J Microbiol Biol Educ 17:156–162. https://doi.org/10.1128/jmbe.v17i1.1011
39. Wooten MM, Coble K, Puckett AW, Rector T. 2018. Investigating introductory astronomy students' perceived impacts from participation in course-based undergraduate research experiences. Phys. Rev. Phys. Educ. Res 14:010151. https://doi.org/10.1103/PhysRevPhysEducRes.14.010151
40. Zelaya AJ, Gerardo NM, Blumer LS, Beck CW. 2020. The bean beetle microbiome project: a course-based undergraduate research experience in microbiology. Front Microbiol 11:577621. https://doi.org/10.3389/fmicb.2020.577621

41. Irby SM, Pelaez NJ, Anderson TR. 2020. Student perceptions of their gains in course-based undergraduate research abilities identified as the anticipated learning outcomes for a biochemistry CURE. J. Chem. Educ 97:56–65. https://doi.org/10.1021/acs.jchemed.9b00440
42. Croonquist P, Falkenberg V, Minkovsky N, Sawa A, Skerritt M, Sustacek MK, Diotti R, Aragon AD, Mans T, Sherr GL, Ward C, Hall-Woods M, Goodman AL, Reed LK, Lopatto D. 2023. The genomics education partnership: first findings on genomics research in community colleges. SPUR 6:17–28. https://doi.org/10.18833/spur/6/3/1
43. Lopatto D, Rosenwald AG, DiAngelo JR, Hark AT, Skerritt M, Wawersik M, Allen AK, Alvarez C, Anderson S, Arrigo C, et al. 2020. Facilitating growth through frustration: using genomics research in a course-based undergraduate research experience. J Microbiol Biol Educ 21:10. https://doi.org/10.1128/jmbe.v21i1.2005
44. 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. https://doi.org/10.1187/09-11-0087
45. Jordan TC, Burnett SH, Carson S, Caruso SM, Clase K, DeJong RJ, Dennehy JJ, Denver DR, Dunbar D, Elgin SC, et al. 2014. A broadly Implementable research course in phage discovery and genomics for first-year undergraduate students. mBio 5:e01051–13. https://doi.org/10.1128/mbio.01051-13
46. Hanauer DI, Graham MJ, Betancur L, Bobrownicki A, Cresawn SG, Garlena RA, Jacobs-Sera D, Kaufmann N, Pope WH, Russell DA, Jacobs WR, Sivanathan V, Asai DJ, Hatfull GF, SEA-PHAGES. 2017. An inclusive research education community (iREC): impact of the SEA-PHAGES program on research outcomes and student learning. Proc Natl Acad Sci U S A 114:13531–13536. https://doi.org/10.1073/pnas.1718188115
47. Vater A, Mayoral J, Nunez-Castilla J, Labonte JW, Briggs LA, Gray JJ, Makarevitch I, Rumjahn SM, Siegel JB. 2021. Development of a broadly accessible, computationally guided biochemistry course-based undergraduate research experience. J. Chem. Educ 98:400–409. https://doi.org/10.1021/acs.jchemed.0c01073
48. Hiatt AC, Hove AA, Ward JR, Ventura L, Neufeld HS, Boyd AE, Clarke HD, Horton JL, Murrell ZE. 2021. Authentic research in the classroom increases appreciation for plants in undergraduate biology students. Integr Comp Biol 61:969–980. https://doi.org/10.1093/icb/icab089
49. 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. https://doi.org/10.1187/cbe-13-08-0152
50. Mathur V, Arora GS, McWilliams M, Russell J, Rosenwald AG. 2019. The genome solver project: faculty training and student performance gains in bioinformatics. J Microbiol Biol Educ 20:20.1.4. https://doi.org/10.1128/jmbe.v20i1.1607
51. Bueso-Bordils JI, Suay-García B, Galiana-Roselló C, Marco-Crespo E, Pérez-Gracia M-T. 2020. Evaluation of the impact of the tiny earth project on the knowledge about antibiotics of pre-university students in the province of valencia on three different school years (2017-2020). Front Microbiol 11:576315. https://doi.org/10.3389/fmicb.2020.576315
52. Lopatto D, Rosenwald AG, Burgess RC, Silver Key C, Van Stry M, Wawersik M, DiAngelo JR, Hark AT, Skerritt M, Allen AK, et al. 2022. Student attitudes contribute to the effectiveness of a genomics CURE. J Microbiol Biol Educ 23:e00208–21. https://doi.org/10.1128/jmbe.00208-21
53. Anderson LJ, Dosch JJ, Lindquist ES, McCay TS, Machado J-L, Kuers K, Gartner TB, Shea KL, Mankiewicz C, Rodgers VL, et al. 2020. Assessment of student learning in undergraduate courses with collaborative projects from the ecological research as education network (EREN). SPUR 4:15–29. https://doi.org/10.18833/spur/4/1/2
54. Rodenbusch SE, Hernandez PR, Simmons SL, Dolan EL, Knight J. 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
55. Vora NJ, Vatcheva K, Saldivar MG, Nair S, Lehker MW, Chew SA. 2023. Biomedical freshman research initiative: A course-based undergraduate research experience at a Hispanic-serving institution. J. Lat. Educ. 22:357–370. https://doi.org/10.1080/15348431.2020.1763352
56. Hewlett JA. 2018. Broadening participation in undergraduate research experiences (UREs): the expanding role of the community college. CBE Life Sci Educ 17:es9. https://doi.org/10.1187/cbe.17-11-0238
57. Leonetti CT, Lindberg H, Schwake DO, Cotter RL. 2023. A call to assess the impacts of course-based undergraduate research experiences for career and technical education allied health, and underrepresented students at community colleges. CBE Life Sci Educ 22:ar4. https://doi.org/10.1187/cbe.21-11-0318
58. Brownell SE, Freeman S, Wenderoth MP, Crowe AJ, Wood WB. 2014. BioCore guide: a tool for interpreting the core concepts of vision and change for biology majors. CBE Life Sci Educ 13:200–211. https://doi.org/10.1187/cbe.13-12-0233
59. Dasgupta AP, Anderson TR, Pelaez N, Dolan E. 2014. Development and validation of a rubric for diagnosing students' experimental design knowledge and difficulties. CBE Life Sci Educ 13:265–284. https://doi.org/10.1187/cbe.13-09-0192
60. Deane T, Nomme K, Jeffery E, Pollock C, Birol G, Wenderoth MP. 2014. Development of the biological experimental design concept inventory (BEDCI). CBE Life Sci Educ 13:540–551. https://doi.org/10.1187/cbe.13-11-0218
61. Feldon DF, Peugh J, Timmerman BE, Maher MA, Hurst M, Strickland D, Gilmore JA, Stiegelmeyer C. 2011. Graduate students' teaching experiences improve their methodological research skills. Science 333:1037–1039. https://doi.org/10.1126/science.1204109
62. Sirum K 2011. The experimental design ability test (EDAT). Bioscene: Journal of College Biology Teaching 37:8–16.
63. Wiersma W. n.d. Educational measurement and testing. Allyn and Bacon, London.
64. Bass KM, Drits-Esser D, Stark LA. 2016. A primer for developing measures of science content knowledge for small-scale research and instructional use. CBE Life Sci Educ 15:rm2. https://doi.org/10.1187/cbe.15-07-0142
65. Love J, Selker R, Marsman M, Jamil T, Dropmann D, Verhagen J, Ly A, Gronau QF, Smíra M, Epskamp S, Matzke D, Wild A, Knight P, Rouder JN, Morey RD, Wagenmakers EJ. 2019. JASP: graphical statistical software for common statistical designs. J Stat Soft 88:1–17. https://doi.org/10.18637/jss.v088.i02
66. Clemmons AW, Timbrook J, Herron JC, Crowe AJ. 2020. Bioskills guide: Development and national validation of a tool for interpreting the vision and change core Competencies. CBE Life Sci Educ 19:ar53. https://doi.org/10.1187/cbe.19-11-0259
67. Spitzer M, Wildenhain J, Rappsilber J, Tyers M. 2014. Boxplotr: a web tool for generation of box plots. Nat Methods 11:121–122. https://doi.org/10.1038/nmeth.2811
68. Craig PA. 2017. A survey on faculty perspectives on the transition to a biochemistry course-based undergraduate research experience laboratory. Biochem Mol Biol Educ 45:426–436. https://doi.org/10.1002/bmb.21060
69. Buonaccorsi VP, Boyle MD, Grove D, Praul C, Sakk E, Stuart A, Tobin T, Hosler J, Carney SL, Engle MJ, Overton BE, Newman JD, Pizzorno M, Powell JR, Trun N. 2011. GCAT-seekquence: genome consortium for active teaching of undergraduates through increased faculty access to next-generation sequencing data. CBE Life Sci Educ 10:342–345. https://doi.org/10.1187/cbe.11-08-0065
70. Bell J, Provost J, Bell E. 2020. The evolution of the malate dehydrogenase cures community. The FASEB Journal 34:1–1. https://doi.org/10.1096/fasebj.2020.34.s1.05443
71. Hurley A, Chevrette MG, Acharya DD, Lozano GL, Garavito M, Heinritz J, Balderrama L, Beebe M, DenHartog ML, Corinaldi K, et al. 2021. Tiny earth: a big idea for STEM education and antibiotic discovery. mBio 12:e03432-20. https://doi.org/10.1128/mBio.03432-20
72. Pieczynski J, Wolyniak M, Pattison D, Hoage T, Carter D, Olson S, Santisteban M, Ruppel N, Challa A. 2022. The CRISPR in the classroom network: a support system for instructors to bring gene editing technology to the undergraduate classroom. The FASEB Journal 36. https://doi.org/10.1096/fasebj.2022.36.S1.R3488
73. Merkle JA, Devergne O, Kelly SM, Croonquist PA, Evans CJ, Hwalek MA, Straub VL, Hamill DR, Peister A, Puthoff DP, Saville KJ, Siders JL, Villanueva Gonzalez ZJ, Wittke-Thompson JK, Bieser KL, Stamm J, Vrailas-Mortimer AD, Kagey JD. 2023. Fly-CURE, a multi-institutional CURE using Drosophila, increases students' confidence, sense of belonging, and

persistence in research. J Microbiol Bio Educ:e00245–22. https://doi.org/10.1128/jmbe.00245-22
74. Staub NL, Poxleitner M, Braley A, Smith-Flores H, Pribbenow CM, Jaworski L, Lopatto D, Anders KR. 2016. Scaling up: adapting a phage-hunting course to increase participation of first-year students in research. CBE Life Sci Educ 15:ar13. https://doi.org/10.1187/cbe.15-10-0211
75. Estrada M, Hernandez PR, Schultz PW, Herrera J. 2018. A longitudinal study of how quality mentorship and research experience integrate underrepresented minorities into STEM careers. CBE Life Sci Educ 17:ar9. https://doi.org/10.1187/cbe.17-04-0066
76. Govindan B, Pickett S, Riggs B. 2020. Fear of the CURE: a beginner's guide to overcoming barriers in creating a course-based undergraduate research experience. J Microbiol Biol Educ 21:50. https://doi.org/10.1128/jmbe.v21i2.2109