External Collaboration Results in Student Learning Gains and Positive STEM Attitudes in CUREs

Kevin P. Callahan,[†]* Celeste N. Peterson,[‡]* Betsy M. Martinez-Vaz[§], Kathryn L. Huisinga[‖], Nicole Galport[¶], Courtney Koletar[¶], Rebecca M. Eddy[¶], Joseph J. Provost[#], Jessica K. Bell[#], and Ellis Bell[#]

[†]Department of Biochemistry, St John Fisher University, Rochester, NY 14618; [‡]Department of Biology, Suffolk University, Boston MA 02108; [§]Department of Biology, Hamline University, Saint Paul, MN, 55104; [‖]Department of Natural Sciences, Malone University, Canton, OH, 44709; [¶]Cobblestone Applied Research & Evaluation, Inc., La Verne, CA 91750; [#]Department of Chemistry and Biochemistry, San Diego, CA, 92110

ABSTRACT

The implementation of course-based undergraduate research experiences (CUREs) has made it possible to expose large undergraduate populations to research experiences. For these research experiences to be authentic, they should reflect the increasingly collaborative nature of research. While some CUREs have expanded, involving multiple schools across the nation, it is still unclear how a structured extramural collaboration between students and faculty from an outside institution affects student outcomes. In this study, we established three cohorts of students: 1) no-CURE, 2) single-institution CURE (CURE), and 3) external collaborative CURE (ec-CURE), and assessed academic and attitudinal outcomes. The ec-CURE differs from a regular CURE in that students work with faculty member from an external institution to refine their hypotheses and discuss their data. The sharing of ideas, data, and materials with an external faculty member allowed students to experience a level of collaboration not typically found in an undergraduate setting. Students in the ec-CURE had the greatest gains in experimental design; self-reported course benefits; scientific skills; and science, technology, engineering, and mathematics (STEM) importance. Importantly this study occurred in a diverse community of STEM disciplinary faculty from 2- and 4-year institutions, illustrating that exposing students to structured external collaboration is both feasible and beneficial to student learning.

INTRODUCTION

Collaboration is essential in the current research environment. Over the last 30 years, the number of collaborative scientific papers has increased dramatically, while the number of single-author papers has dropped (Jones *et al.*, 2008). This shift toward collaboration is most likely because modern research requires extensive specialization; scientists today are expected to be heavily specialized in one field while being conversant in other fields (Kuczenski *et al.*, 2005; Labov *et al.*, 2010; National Research Council, 2012). When identifying collaborators, researchers increasingly pay less attention to location and instead focus on expertise, which is often found at other institutions. Over the last few decades, external collaborations have grown from accounting for no more than 10% in the 1970s to greater than 35% in 2005 (Jones *et al.*, 2008). Technological advances and increased familiarity with meeting virtually assures that this type of extramural collaborative research will continue to grow. Furthermore, external collaborations are also more likely to produce higher-impact papers than single-institution collaborations (Jones *et al.*, 2008). As undergraduate educators interested in course-based undergraduate research experiences (CUREs), we aim to give our students the most realistic research experience, including exposure to this

Lisa Corwin, *Monitoring Editor*

Submitted Jul 1, 2021; Revised Jul 26, 2022; Accepted Aug 16, 2022

CBE Life Sci Educ December 1, 2022 21:ar74

DOI:10.1187/cbe.21-06-0167

[†]These authors contributed equally to this work.
*Address correspondence to: Kevin P. Callahan (kcallahan@sjfc.edu) or Celeste N. Peterson (cnpeterson@suffolk.edu).

type of external collaboration that frequently occurs in the research world.

Many schools have implemented CUREs to increase access to research opportunities for undergraduates. These high-impact pedagogical approaches teach students critical skills with five different elements: 1) the use of scientific practice, 2) discovery, 3) broadly relevant/important work, 4) collaboration, and 5) iteration (Auchincloss *et al.*, 2014). Collaboration has long been known to increase student performance and motivation in the classroom, including science classes (Linton *et al.*, 2014; Scager *et al.*, 2016). In the context of CUREs, collaboration and iteration have been shown to be key to students' sense of ownership of their course work (Corwin *et al.*, 2015b, 2018). A growing number of CUREs encompassing a series of linked courses within the same department or school have been described (Luckie *et al.*, 2004; Kowalski *et al.*, 2016). However, nearly all studies on CURE collaboration have been involved students within the same course or institution. We are not aware of any studies that have addressed extramural collaboration between students and faculty.

Over the last decade, the number of CUREs formed with a consortium of schools has increased. These research education communities—SEA-PHAGES, Genomics Education Partnership, and Small World Initiative—have been successful in promoting undergraduates from many different schools to work on similar projects, and sometimes even include interactive symposia for the students (Shaffer *et al.*, 2010; Caruso *et al.*, 2016; Hanauer *et al.*, 2017). In addition, several consortia such as BASIL and iCURE promote research aimed at upper-level students (Roberts *et al.*, 2019; Stoeckman *et al.*, 2019). All of the consortium CUREs have demonstrated the benefits of pooling resources, both intellectual and physical. However, these consortia did not formally establish direct collaborations between students and a faculty member from a different institution.

The Malate Dehydrogenase CUREs Community (MCC) consists of a diverse community of science, technology, engineering, and mathematics (STEM) disciplinary faculty members from institutions that vary across a number of dimensions, including type (2-year vs. 4-year), enrollment size, selectivity, and student population (low income, first generation). Students enrolled in the CURE use facets of bioinformatics, 3D structure visualization, and pertinent background literature to construct a novel hypothesis about the role of a specific amino acid in the activity of malate dehydrogenase (MDH).

One aim of MCC was to determine whether the incorporation of structured external collaborations in CUREs enhances student learning outcomes. To date, only a few courses have adopted formal multi-institution collaborations (Knisley and Behravesh, 2010) and their effect on student learning is not well understood. We hypothesized that collaboration with a faculty member from an external institution would result in higher student learning gains and engagement with the course materials. We established three cohorts of students across a range of schools: 1) no CURE, 2) single-institution CURE (CURE), and 3) external collaborative (ec-CURE); and we assessed their academic and attitudinal outcomes over the course of the semester (Table 1). Students in CURE worked in teams within their course but did not collaborate with other institutions. Students in ec-CURE worked in teams within their course and also collaborated with faculty from an external institution at specific times during the semester (Table 2). To examine the benefits of collaboration on students' science skills and attitudes, students in each cohort completed a pre–post Experimental Design Ability Test (EDAT), pre–post Test of Scientific Literacy Skills (TOSLS), and post surveys measuring their learning gains and attitudes (Sirum and Humburg, 2011; Gormally *et al.*, 2012). Students in the ec-CURE had the greatest gains in experimental design, self-reported course benefits, self-reported skills, and STEM importance. This study indicates that it is possible to expose undergraduates to a structured external collaboration in the classroom setting, and this experience is beneficial to students at a wide range of schools.

METHODS

Study Context and Experimental Design

The MCC was established to develop a sustainable protein-centric CURE and investigate critical questions on undergraduate pedagogy. A central question of MCC was to determine the effect of collaboration, specifically structured external collaboration, on student learning. Not only is it important to collaborate with peers within a laboratory setting (the common structure for CUREs), but in today's interdisciplinary world, it is vital to collaborate with other research groups and institutions. There is little research, however, that examines the impact of cross-institutional CURE collaboration.

Beginning in the Fall of 2017 and concluding in Spring 2020, a diverse community of faculty and students participated in the MCC. Institutions varied across a number of dimensions including type (2-year vs. 4-year), enrollment size, selectivity, and student population (low income, first generation; Tables 1 and 3 and Supplemental Material 1). Courses included first-year courses, sophomore and junior year biochemistry and molecular

TABLE 1. Breakdown of Institutions and students enrolled in the study. The percentage represents the percent of students from the respective institution type (PUI, R1, CC) within each experimental group (No-CURE, CURE, ec-CURE)

	Primarily undergraduate	Research intensive	Community college
No-CURE			
courses	12	12	6
Students	227 (38%)	253 (42%)	123 (20%)
CURE			
courses	14	6	5
Students	175 (30%)	216 (37%)	194 (33%)
ec-CURE			
courses	12	4	3
Students	160 (55%)	101 (35%)	29 (10%)

TABLE 2. Comparison of the No CURE, CURE and ec-CURE cohorts in the study

	No CURE	CURE	ec-CURE
Introduction to project	✓	✓	✓
Read background literature and develop novel hypothesis		✓	✓
Collaborate with home institution faculty and students	✓	✓	✓
Collaborate with faculty from external institution			✓
Perform experiments and data analysis	✓	✓	✓
Evaluate results and conclusions with faculty and students from home institution	✓	✓	✓
Evaluate results and conclusions with faculty and students from external institution			✓
Present results in written or oral format	✓	✓	✓

biology courses, and more specialized senior-level capstone courses. The broad range of faculty and students participating in the MCC is a strength of the community but also presents some challenges in terms of data analysis. As discussed in the *Limitations* section, the assignment of ec-CURE, CURE, and no-CURE courses was not always random and resulted in an uneven distribution of course conditions by institution type and course level (i.e., lower division vs. upper division). However, it is noteworthy that, over the course of the study, all of the faculty who taught an ec-CURE also taught in CURE and/or no-CURE groups that were included in the study.

To determine the impact of extramural collaboration, we evaluated students in three different cohorts: 1) no CURE, 2) single-institution CURE (CURE), and 3) external collaborative CURE (ec-CURE). To control for differences between groups, students in all cohorts learned laboratory techniques, performed data analysis, reported their results either orally or in writing, and collaborated with faculty mentors and students from their home institutions. In addition, the no-CURE, CURE, and ec-CURE cohorts uniformly engaged in collaborative/evaluative/reflective features of the lab. A variable between the cohorts was students in the CURE and ec-CURE developed and interrogated a novel hypothesis around a specific research project involving MDH, whereas the no-CURE students did not develop a hypothesis and did not always work on MDH. The most critical variable was that only students in the ec-CURE collaborated with faculty from an external institution (Table 2). These students in the ec-CURE sections formally met at least twice with a faculty from an outside institution with significant expertise working on MDH. Meetings typically took place virtually and were facilitated by the students' home institutions during their scheduled lab sections. Before the first meeting, students read background literature, developed their own hypotheses, and prepared slides to share with their collaborators. In the meeting, the external collaborator gave feedback on the hypothesis, and sometimes a second meeting was scheduled to discuss a revised hypothesis. Toward the end of the semester, a final meeting was arranged. The students presented their data and discussed whether the data supported/refuted their respective hypotheses. The students and the external collaborator talked about how the results connected to other research in the field, discussed any pitfalls, and suggested new models and ways to test them. To help enhance the experience of the students, we developed guidelines and best practices for these collaborative sessions (Supplemental Material 2).

TABLE 3. Student demographics

	No CURE	Ec-CURE	CURE	Total
Number of students	603 (100%)	290 (100%)	585 (100%)	1478 (100%)
Gender				
Woman	371 (61.5%)	168 (57.9%)	331 (56.6%)	870 (58.9%)
Man	230 (38.1%)	122 (42.1%)	244 (41.7%)	596 (40.3%)
Unknown[a]	2 (0.3%)	0 (0.0%)	10 (1.7%)	12 (0.8%)
Race/ethnicity				
White/Asian	418 (69.3%)	229 (79.0%)	338 (57.8%)	985 (66.6%)
URM[b]	145 (24.0%)	41 (14.1%)	195 (33.3%)	381 (25.8%)
Other[c]	40 (6.6%)	20 (6.9%)	52 (8.9%)	112 (7.6%)
Student level				
First/second year	182 (30.2%)	26 (9.0%)	141 (24.1%)	349 (23.6%)
Third/fourth year	373 (61.9%)	255 (88.0%)	410 (70.0%)	1038 (70.2%)
Other[d]	48 (8.0%)	9 (3.1%)	34 (5.8%)	91 (6.1%)
Institution type				
CC	123 (20.4%)	29 (10%)	194 (33.1%)	346 (23.4%)
PUI	227 (37.6%)	160 (55.1%)	175 (30.0%)	562 (38.0%)
RI	253 (42.0%)	101 (34.8%)	216 (37.0%)	570 (38.5%)

CC: Community College; PUI: Primary Undergraduate; RI: Research Intensive.
[a]Missing or prefer not to answer.
[b]American Indian or Alaska Native, Black or African American, Native Hawaiian or other Pacific Islander, two or more races.
[c]International students and unknown.
[d]Postbaccalaureate, high school, and unknown.

Collaboration took place on an intellectual level, with students and external faculty sharing ideas, and some collaborations also involved the exchange of physical materials. For example, students at Southwest Community College collaborated with a professor from the University of San Diego (USD) in a CURE focusing on intramolecular forces in which they performed docking predictions between cytosolic MDH and several other proteins. The external professor "visited" the class several times to present the background science and interact with the students on their ideas/hypotheses of the interactions and then joined two other USD professors to form a review panel for students' final presentations. In another example, faculty, from St. John Fisher College (SJFC) and the USD were teaching CUREs that employed site-directed mutagenesis to study the catalytic mechanism of MDH. The faculty shared their expertise with the other institutions' students as students were developing their respective hypotheses. Later in the semester, the students exchanged purified mutant proteins for further analysis; circular dichroism was performed at USD and native gel analysis was performed at SJFC. The students were able to experience firsthand both the intellectual and physical advantages of real collaboration.

Institutional Review Board Approval and Data Collection

The MCC evaluation received institutional review board approval for all participating institutions before data collection. Instruments included the TOSLS (Gormally *et al.*, 2012), the EDAT (Sirum and Humburg, 2011), CURE pre–post survey (Lopatto *et al.*, 2008), and STEM Career Interest Questionnaire (Tyler-Wood *et al.*, 2010). Students were provided a link to complete the study instruments; consent forms and pretest assessments were administered to students on the first day of class, and posttest assessments were administered on the last day of class. All evaluation surveys were completed online, with the exception of the EDAT, which students completed in pen-and-paper format (Sirum and Humburg, 2011).

Data Analysis

Assessment data were analyzed at the conclusion of data collection in 2020. Analyses of covariance (ANCOVAs) were used to compare differences in EDAT scores, TOSLS scores, science attitudes, and learning gains at posttest while controlling for pretest responses. This analysis approach was selected to compare if and why there might be differences between the three conditions on posttest scores (Wright, 2006). In addition, pretest response scales differed from posttest response scales for the science attitudes and learning gains items, which does not allow for a direct comparison on gains from pre to post. The assumptions of the ANCOVA were tested by examining the skewness (normality), using a Levene's test (homogeneity of variances), and regressing each condition, the independent variable, and the interaction between each condition and the independent variable on the dependent variable (homogeneity of regression slopes). Skewness is reported as follows: $< |1|$ as normally distributed, $> |1|$ and $< |1.5|$ as slightly skewed, $> |1.5|$ and $< |2|$ as moderately skewed, and $> |2|$ as highly skewed. Of note, none of the data reached the threshold of highly skewed. All data for ANCOVAs also met the homogeneity of regression slopes assumption. Some data demonstrated slight to moderate nonnormality and/or heterogeneity of variances, and this is noted within the *Results* section for transparency. However, the ANCOVA is robust to the violation of the normality and homogeneity of variances assumptions when sample sizes are large enough (Glass *et al.*, 1972; Knief and Forstmeier, 2021). One-way analyses of variance (ANOVAs) were used to compare across the three conditions for posttest-only measures, and Tukey post hoc tests were used to explicate significant findings. The assumptions of the ANOVA (i.e., normality, homogeneity of variances) were tested in the same ways as the same assumptions of the ANCOVA. Similarly, some data demonstrated slight to moderate nonnormality and this is noted in the *Results* section for transparency. However, the ANOVA is robust to the violation of the normality assumption when sample sizes are large and the homogeneity of variances assumption is met (Olegjnik and Algina, 1983). All information identifying students was redacted to reduce coder bias. To maintain the confidentiality of all participants, data were analyzed and presented in aggregate, and narrative accounts were de-identified.

Laboratory Course Assessment Survey (LCAS)

The LCAS measures student perceptions of three aspects of CUREs: collaboration, discovery and relevance, and iteration (Corwin *et al.*, 2015b). Students reported how frequently six aspects of collaboration occurred on a scale from 1 = "never" to 4 = "weekly" and rated their level of agreement with five statements about discovery and six statements about iteration on a scale from 1 = "strongly disagree" to 7 = "strongly agree." A total of 1143 students responded to the survey after completing their ec-CURE (n = 260), CURE (n = 426), or no-CURE (n = 457) course. Composite mean scores were calculated for two of the three CURE components (discovery and relevance and iteration). Collaboration was rated on an ordinal scale and was therefore analyzed on an item-by-item basis using nonparametric tests.

Kruskal-Wallis tests were conducted to compare answers on the six collaboration scale items between conditions.

For discovery and relevance, distributions were negatively skewed, and variances were unequal across conditions. Both one-way ANOVA and nonparametric Kruskal-Wallis tests were run to compare results between the two methods. Results did not differ between analysis methods, and one-way ANOVA results are reported here. There was a significant difference in discovery and relevance between conditions, $F(2, 1131) = 114.32$, $p < 0.001$. Tukey post hoc tests found greater discovery and relevance in ec-CURE and CURE courses than no-CURE courses ($p < 0.001$ for both).

TOSLS

A total of 1115 students completed both the pretest and the posttest. An ANCOVA was used to examine differences across groups on posttest TOSLS scores while controlling for pretest TOSLS scores. The data demonstrated homogeneity of variances and regression slopes. The covariate had a moderately nonnormal distribution for the collaborative and independent conditions; however, the ANCOVA test remains robust to a violation of the conditional normality assumption when there is homogeneity of variances (Olejnik and Algina, 1983) and regression slopes (Levy, 1980). There was no significant difference in posttest scores across conditions, $F(2, 1111) = 1.26$, $p < 0.05$, meaning that TOSLS scores did not differ across ec-CURE, CURE, and no-CURE courses.

EDAT

The EDAT measures student understanding of the criteria for experimental design. The EDAT asks respondents to describe the experiment they would design in response to an everyday life science prompt. A training session was held in August 2017 with 14 of the 15 participating faculty members for the purpose of establishing interrater reliability for scoring EDATs. During the session, faculty members read three examples and provided an independent rating of each sample on 10 dimensions (using the standard EDAT scoring rubric); the research team calculated the level of agreement for each of the dimensions. During the session, sample entries were used as a basis for discussion when disagreements in ratings were observed. These discussions further established rules for future ratings. In 2018, a second session was held that incorporated new faculty members into the process for the purpose of once again establishing interrater reliability.

A total of 1187 students completed the EDAT both before and after completing the course. Each EDAT was scored by two faculty members from institutions other than the students' own. Faculty members indicated the presence or absence of 10 elements of experimental design, resulting in scores ranging from 0 to 10. Some completed EDAT posttests in year 3 were not rated by any faculty members, bringing the total matched pretest and posttest number to 1139. Scores from three raters who had an interrater reliability of less than 0.500 on the posttest were removed, resulting in 1044 matched student pretest and posttest scores available for analysis. For EDAT responses with two usable ratings, the student's score was represented by the mean of the two ratings. For EDAT responses with only one usable rating, the single rater's rating was used as the student's score. For the 862 pretest responses with valid scores from two faculty members, there was an interrater reliability of 0.704 ($p < 0.001$). For the 629 posttest responses with valid scores from two faculty members, there was an interrater reliability of 0.666 ($p < 0.001$). These correlations are below those reported in the literature ($r = 0.835$, $p < 0.001$).

CURE Pre and Post Survey on Learning Gains, Science Attitudes, Course Benefits, and Overall Student Evaluations

For the learning gains pretest, students rated their experience with 25 activities that occur in science courses using a scale from 1 = "No experience or feel inexperienced" to 5 = "Extensive experience or mastered this element." For the learning gains posttest, students rated their learning gains related to each of the 25 activities using the scale 1 = "No gain or very small gain" to 5 = "Very large gain." The items were found to be internally consistent at both pretest ($\alpha = 0.91$) and posttest ($\alpha = 0.93$), and a mean composite was calculated for each before carrying out a series of ANCOVAs across the conditions while controlling for the pretest response. Because 25 tests were run, the critical α value was adjusted using a Bonferroni correction to $p = 0.002$ (0.05/25). Twelve of the learning gains posttest items were moderately negatively skewed. This may have slightly reduced the power of the corresponding F-tests (Glass *et al.*, 1972) Similarly, six items had heterogenous variances. For all six items, the degree of variance heterogeneity was small (i.e., variance ratios less than 1:1.3).

On the CURE pretest, students responded to 22 items concerning their attitudes toward science using a scale from 1 = "strongly disagree" to 7 = "strongly agree." On the posttest, students responded to the same 22 items using a scale from 1 = "strongly disagree" to 5 = "strongly agree." Five of the items comprised a positive attitudes toward science scale: The inter-item correlations were acceptable at both pretest ($\alpha = 0.71$) and posttest ($\alpha = 0.75$), and mean composite scores were calculated for each. Six of the items comprised a negative attitudes toward science scale: The inter-item correlations were low at pretest ($\alpha = 0.65$) and acceptable at posttest ($\alpha = 0.80$). Mean composite scores were calculated for each; however, results should be interpreted with caution. The remaining 11 items were analyzed individually, and the critical α values were adjusted using a Bonferroni correction to $p = 0.005$ (0.05/11). The positive attitudes toward science scale, the negative attitudes toward science scale, and the individual items were all analyzed using ANCOVAs that examined differences across conditions at posttest while controlling for pretest responses. Six of the individual science attitudes items as well as the positive attitudes toward science composite were moderately skewed. One individual item and the positive attitude toward science composite had heterogenous variances; however, variance ratios were small (i.e., less than 1:1.3).

On the CURE posttest survey, students were asked to report on the benefits of their course using a scale from 1 = "No gain or very small gain" to 5 = "Very large gain." The items demonstrated high internal consistency ($\alpha = 0.97$), and a mean composite was calculated.

On the CURE posttest survey, students responded to four items evaluating their course as a whole. Students rated their level of agreement with each item on a scale from 1 = "strongly disagree" to 7 = "strongly agree." Items were analyzed individually.

STEM Career Interest

After completing the course, students rated their agreement with 12 statements concerning their interest in STEM on a scale from 1 = "strongly disagree" to 5 = "strongly agree." The statements were grouped into three subscales composed of four statements each: STEM Support, STEM Career Interest, and STEM Importance. Tyler-Wood *et al.* (2010) describe STEM support as "perception of supportive environment for pursuing a career in science"; STEM career interest is described as "interest in pursuing educational opportunities that would lead to a career in science"; and STEM importance is described as "perceived importance of a career in science" (p. 351). A composite mean was calculated for each of the three subscales ($\alpha = 0.83$–0.89).

Faculty Focus Groups and Survey

Faculty focus groups were administered by a member of the Cobblestone evaluation team at the end of each year for years 1, 2, and 3 of the grant. Participating faculty were asked to reflect on developing, implementing, and improving their CURE courses. Faculty members were also be asked about the extent to which they collaborated with other faculty members.

To further investigate the faculty MCC experiences, a survey was disseminated to all faculty members of the community at the conclusion of the grant. The survey focused on two components of CURE implementation: hypothesis development (as part of the scientific process) and collaboration. All faculty were asked four questions to better explore their experiences of collaboration as part of implementing a CURE. They rated their level of agreement with each item on a scale from 0 = "not at

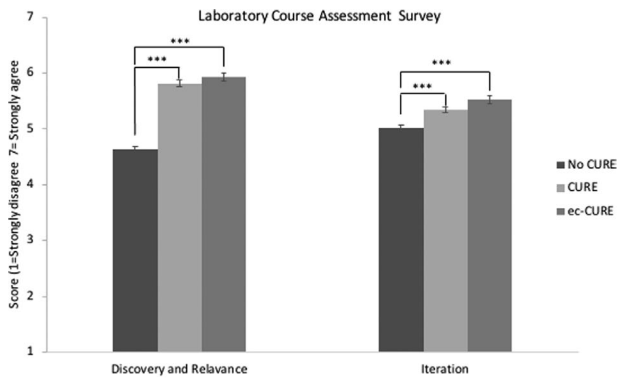

FIGURE 1. Student perceptions of two aspects of CURE—discovery and relevance, iteration—as measured by the LCAS (Corwin *et al.*, 2015b). To compare differences across conditions (no CURE, CURE, ec-CURE), we conducted a series of one-way ANOVAs. There was a significant difference in discovery and relevance between conditions, $F(2, 1131) = 114.32$. Tukey post hoc tests found that there was greater discovery and relevance in ec-CURE and CURE courses than in no-CURE courses. ***$p < 0.001$.

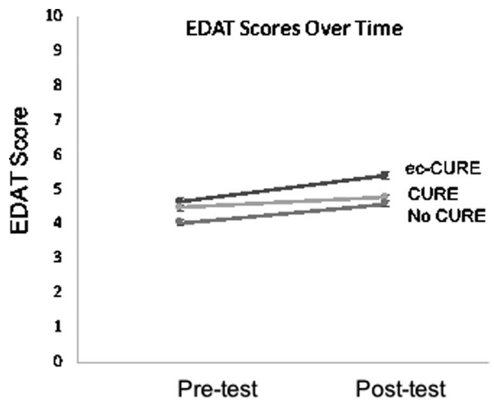

FIGURE 2. Comparison of pre- and posttest EDAT scores, which report students' experimental design knowledge. A total of 397 students in the no-CURE courses, 387 students in CURE courses, and 260 in ec-CURE had matched ratings for both EDAT tests. The ANCOVA controlling for pretest EDAT scores found a significant difference in posttest EDAT scores between the three types of courses ($p < 0.001$). Comparison of the estimated marginal means using Fischer's least significant difference indicated that students in the ec-CUREs had higher EDAT scores at posttest than students in the no-CURE ($p < 0.001$) and CURE ($p < 0.001$) courses.

all important" to 3 = "critically important" and were asked to provide a reason or reasons for their respective rating (Supplemental Material 3). All faculty signed a consent form, and the data from the survey were collected and tabulated anonymously using the Qualtrics platform. The survey was approved by the USD Institutional Review Board.

RESULTS

Students Perceptions of Laboratory Course Design

The LCAS measures student perceptions of three aspects of CUREs: collaboration, discovery and relevance, and iteration. It has been found that the LCAS can differentiate between CUREs and traditional laboratory courses on the discovery and relevance, and iteration scales (Corwin *et al.*, 2015b). There was a significant difference in discovery and relevance between conditions, $F(2, 1131) = 114.32$, $p < 0.001$. Tukey post hoc tests found that there was greater discovery and relevance in ec-CURE and CURE courses than no-CURE courses ($p < 0.001$ for both; Figure 1). Thus, our no-CURE courses and our CURE courses were consistent with previous studies of CURE design.

The LCAS also examines the degree of collaboration perceived by the students during the course. Students reported how frequently six aspects of collaboration occurred on a scale from 1 = "never" to 4 = "weekly" and rated their level of agreement with five statements about discovery and six statements about iteration on a scale from 1 = "strongly disagree" to 7 = "strongly agree." This aspect has not been shown to be significant between CURE and no-CURE courses. There were significant differences between conditions on two items: "In this course I was encouraged to discuss elements of my investigation with classmates or instructors," $H(2) = 10.57$, $p < 0.01$, and "In this course I was encouraged to contribute my ideas and suggestions during class discussions," $H(2) = 7.84$, $p < 0.05$. For both items, students in ec-CUREs reported more frequent collaboration than students in CURE and no-CURE courses. While the LCAS instrument did not distinguish between internal and external collaboration, it did demonstrate that students in the ec-CURE were experiencing more collaboration than students in the other courses.

Student Learning Gains

To follow student learning gains in the different cohorts, we employed both the EDAT and the TOSLS. The EDAT measures students' knowledge of experimental design by asking the students to describe the design of an experiment that would test a given everyday claim (Sirum and Humburg, 2011). Students completed the EDAT before and after the course, and their responses were scored based on the presence or absence of 10 factors in experimental design (Sirum and Humburg, 2011). Over 3 years, a total of 414 students in the no-CURE courses, 461 students in CUREs, and 261 students in ec-CUREs completed both the EDAT pretest and posttest. Students in ec-CUREs scored higher on average than students in CUREs and also the no-CURE courses, $F(2, 1133) = 22.95$, $p < 0.001$, partial eta-squared = 0.04. There was also a significant interaction between time and CURE type, $F(2, 1133) = 3.70$, $p < 0.05$, partial eta-squared = 0.01. A Tukey post hoc analysis suggested that participating in an ec-CURE led to more growth over time than a CURE ($p < 0.05$) or no CURE ($p < 0.05$; Figure 2). Students also completed pre and post TOSLS, which is an assessment of students' ability to evaluate scientific information and scientific arguments (Gormally *et al.*, 2012). A mixed-factorial ANOVA showed that there was no significant main effect of condition ($p > 0.05$) and no significant interaction between the time and condition ($p > 0.05$), meaning that TOSLS scores did not differ across ec-CURE, CURE, and no-CURE courses. There was a main effect of time that, surprisingly, found scores for all three groups dropped significantly during the semester, but there was no significant interaction between time and condition (Supplemental Material 4). Thus, engaging in a collaborative CURE was associated with student gains in experimental design but did not have an effect on scientific literacy.

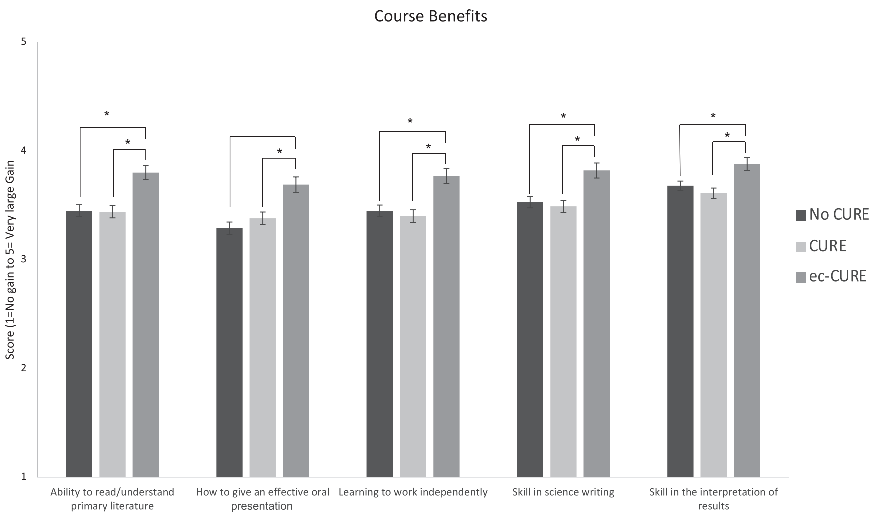

FIGURE 3. Student ratings on course benefits. To compare differences across conditions (no CURE, CURE, ec-CURE) on specific course elements, a series of one-way ANOVAs were conducted. Tukey post hoc tests were performed; effects that were significant are noted on the graph: *$p < 0.02$.

Course Benefits and Overall Course Evaluation

In addition to learning outcomes, students were surveyed about the benefits of the course. After completion of the course, they were asked to reflect on a variety of possible benefits they may have gained from their experience (Lopatto *et al.*, 2008). For 21 possible benefits listed, students were asked to report on a scale of 1 = "No gain or very small gain" to 5 = "Very large gain." There was a significant difference in course benefit scores across CURE type, $F(2, 1120) = 6.14$, $p < 0.01$, partial eta-squared = 0.01. A Tukey post hoc test found that benefit ratings were significantly higher ($p < 0.01$) in the ec-CURE condition (mean = 3.76) than ratings in both the CURE (mean = 3.54) and no-CURE condition (mean = 3.56). There were five reported benefits that helped drive this significant effect ($p < 0.002$). These benefits fall into categories of communication ("skills in how to give an effective oral presentation" and "skills in writing") and helping the students develop into independent scientists (including "ability to read and understand primary literature," "learning to work independently," and "skill in the interpretation of results"; Figure 3). Of the 16 remaining items, 11 had a higher mean in the ec-CURE classes, with a $p < 0.05$ ($F > 3$; Supplemental Material 5). Thus, students perceived that many benefits occurred to a greater extent in the ec-CUREs compared with the CUREs or no-CURE courses.

At the end of the course, students were asked to evaluate the course as a whole based on four statements. They rated their level of agreement with each item on a scale from 1 = "strongly disagree" to 7 = "strongly agree." The distributions for all four items were negatively skewed (Figure 4).

Students' Self-Reported Learning Gains and STEM Career Interest

In an effort to examine the benefits of collaboration on students' self-reported science skills and attitudes, we compared learning gains among participants across 25 elements on the CURE pretest and posttest survey (Lopatto *et al.*, 2008). At pretest, students rated their experience with 25 activities that occur in science courses using a scale from 1 = "No experience or feel inexperienced" to 5 = "Extensive experience or mastered this element." At posttest, students rated their learning gains related to each of the 25 activities using the scale 1 = "No gain or very small gain" to 5 = "Very large gain." Overall, eight of the 25 activities showed significant differences in learning gains between any type of CURE (CURE or ec-CURE) and the no-CURE conditions. There were two activities for which ec-CUREs were significantly different from both the CURE and no-CURE conditions. Post hoc simple contrasts found that learning gains for "A lab or project where no one knows the outcome" were significantly greater in the ec-CURE than the CURE ($p = 0.025$) and the no CURE ($p < 0.001$; Figure 5). In addition, students in the ec-CUREs reported significantly greater gains in their ability to present orally compared with students in the CURE ($p = 0.009$) and no-CURE conditions ($p < 0.001$; Figure 5).

Additionally, students were asked 22 questions about their attitudes about science (Lopatto *et al.*, 2008). The results of the ANCOVA analyses were mixed. Conditions differed significantly on the positive attitudes toward science scale, $F(2,1118) = 5.82$, $p = 0.003$, partial eta-squared = 0.010. Ratings in the ec-CURE ($p = 0.001$) and no-CURE ($p = 0.049$) conditions were unexpectedly significantly greater than ratings in the CURE condition. There were no differences between conditions on the negative attitudes toward science scale or any of the 11 individual science attitudes items.

To investigate the impact of collaboration on students' interest in STEM, students were given a post-course survey

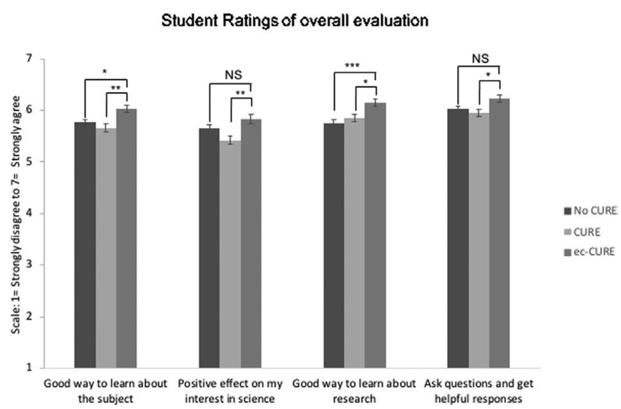

FIGURE 4. Student ratings on overall evaluation. To compare differences across conditions (no CURE, CURE, ec-CURE) on the overall evaluation items, a series of one-way ANOVAs were conducted. Tukey post hoc tests were performed; effects that were significant are noted on the graph: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.

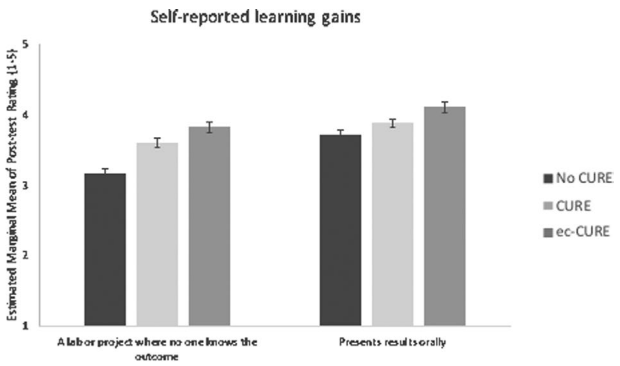

FIGURE 5. Self-reported learning gains. To compare differences across conditions (no CURE, ec-CURE, CURE) on the learning gains items, a series of ANCOVAs controlling for prior experience with each activity were conducted. Simple contrast post hoc tests were performed; effects that were significant are noted on the graph: ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. Figure displays the estimated marginal means for each condition (lab or project where no one knows the outcome evaluated at a pretest value of 2.57; present results orally evaluated at a pretest value of 3.12).

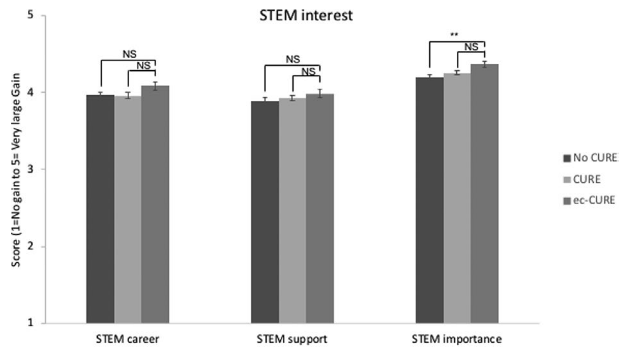

FIGURE 6. STEM interest of students. To compare differences across conditions (no CURE, CURE, ec-CURE) on students' interest and understanding of STEM at the end of the courses, a series of one-way ANOVAs were conducted, and Tukey post hoc tests were performed. Effects that were significant are noted on the graph: **$p < 0.01$.

containing 12 statements about previous research experiences, and a Likert-type STEM career interest scale (Tyler-Wood *et al.*, 2010). The statements in the survey were divided into three categories: STEM support, STEM career interests, and STEM importance. A series of one-way ANOVA tests were conducted to assess differences in students' reported STEM interest across the three groups (no CURE, CURE, ec-CURE). Distributions for all three constructs were negatively skewed.

There was a significant difference between conditions on STEM importance, $F(2, 1119) = 5.35$, $p < 0.01$. Tukey post hoc tests found that students in the ec-CUREs students had significantly higher STEM importance scores compared with participants in the no-CURE group ($p < 0.01$). Assessment of the STEM support and STEM career interests' components of the post-course survey showed greater values in the ec-CURE group compared with the no CURE and CURE; however, these differences were not significant (Figure 6).

Faculty-Reported Student Learning Gains, Student Attitudes, and Faculty Satisfaction

Faculty surveys indicated that the collaboration had a strong effect on student learning and attitudes. Faculty participated in a written survey after the completion of 3 years in the MCC. A strong majority (91%) thought that the collaboration was critically important or important to student learning gains (Table 4). One faculty member commented, "Collaboration with a different institution increased students' learning gains related to data analysis and interpretation. Students were also better at keeping records of their data (notebooks) and communicating their results. They put more effort into these aspects of the project because they want to be good collaborators." Furthermore, 82% percent of the faculty thought that the collaboration was important or critically important to affecting student attitudes toward the scientific process (Table 4). One faculty member commented that the students "better understood that hypothesis development is a very iterative process that requires a significant amount of background information and feedback from other people."

Finally, the survey results indicated that the collaboration also offered the faculty both scientific support and encouragement. A strong majority (91%) of the faculty said that collaboration was critically important or important to their satisfaction doing the CURE (Table 4). In the comments, they specifically commented on how collaboration with an outside expert provided technical support and "increased my understanding of the MDH and the CURE itself." The faculty also noted how working with others was motivating. One person commented, "Having other faculty colleagues provided moral and technical support, was helpful in having someone to think through ideas with and sometimes helped me to get things done because I knew someone else wanted it too."

DISCUSSION

There is significant evidence that undergraduate research experiences have a positive impact on student learning (Auchincloss *et al.*, 2014; Corwin *et al.*, 2015b). The implementation of CUREs has made it possible to expose large undergraduate populations with diverse backgrounds to authentic research experiences, and there is a need to continually enhance this model (Auchincloss *et al.*, 2014; Corwin *et al.*, 2015a, b). One element of the CURE experience that we sought to implement and study was collaboration. Clearly, collaboration takes place within a typical course setting between lab partners, classmates, and an instructor, but does collaboration with a faculty member from outside the home institution enhance the student experience? The MCC has emphasized CUREs that employ multischool collaboration across a diverse set of institutions and courses. In addition to establishing a model for this type of external collaboration (Supplemental Material 2), we found that students performed better on the EDAT test, had better ratings of the course, had better self-reported skills, and saw an increased importance in STEM fields when enrolled in the ec-CURE compared with the no-CURE courses. This suggests that the presence of an external collaborator helped students learn and had positive effects on their attitudes toward science. Finally, most faculty reported that working collaboratively was important to their overall satisfaction.

TABLE 4. Faculty survey results

	Critically important	Important	Somewhat important	Not important
How did collaboration affect student learning gains?	64%	27%	9%	0%
How did collaboration affect student attitudes toward the scientific process?	64%	18%	18%	0%
How did collaboration affect your satisfaction with the course?	55%	36%	9%	0%

External Collaboration Increases Student Learning

The ec-CURE experience had a positive effect on student learning outcomes, as demonstrated by both the EDAT assessment and student-perceived increase in scientific skills (Figures 2 and 3). Students in the ec-CURE increased their EDAT scores (a measure of student understanding of the criteria for good experimental design) more over the course of the semester than students in the CURE or no-CURE courses (Figure 2). Consistent with the increase in EDAT scores, students in the ec-CURE also self-reported greater gains in specific scientific skills such as reading primary literature, interpreting results, and learning associated with "A lab or project where no one knows the outcome" (Figures 3 and 5).

Science is a highly iterative process, and having students explain their projects and their thinking to an external collaborator who is an expert in the field appears to lead to a deeper understanding of the material. MCC faculty participated in focus groups every year to share their experiences. Faculty who participated in ec-CUREs often reflected on how the collaboration task was challenging for the students but ultimately led to growth and helped cement students' understanding of the project. In a focus group, one faculty member said, "The students were scared to present (their project to the collaborator). But then afterwards they said, I see why you made us do it. It solidified (the project) in their head much more than any other assignment I've done." The open-ended and complex nature of CUREs may be especially effective at promoting learning through collaboration, as these types of assignments tend to improve reasoning and evaluative thinking more than closed tasks (Gillies, 2014). Finally, interacting with an external collaborator could lead to better student metacognition through reflecting on their own misconceptions and discussing how the material relates to knowledge that they already have (Tanner, 2012; Wittrock, 1989).

External Collaboration Positively Affects Student Attitudes and Communication Skills

Students in the ec-CUREs also reported a more positive outlook on the course and science in general. Overall, they noted that the ec-CURE courses had greater benefits than the CURE or the no-CURE courses and were more likely to describe this course as a good way to learn about their subject and research. Faculty noted that students accepted the legitimacy of the research when they knew that other institutions were also working on the project. One faculty member said, "Collaborating with an outside institution increased the students' engagement in the project. Knowing that other students/faculty were working on the same enzyme just made what they were doing feel all the more important. Most of my students had never experienced anything like this before in any of their other classes." This effect is consistent with students reporting a more positive effect on their interest in science and an increased importance of STEM (Figure 6). Another faculty member recalled when a student in the ec-CURE group asked: "Are you saying that we've been asked to participate in a national collaboration? Wow, like I've been chosen." Thus, having an external collaborator stimulated student interest in science.

For successful collaborations, strong communications skills are needed. How students communicate with one another and the quality of their interactions—such as how well they can incorporate and build upon the ideas of others—affects their learning outcomes (Chinn *et al.*, 2000; Barron, 2003; Volet *et al.*, 2009). Our data suggest that the ec-CURE experience increased communications skills (oral presentation and writing skills) and also the ability to delve into discourse (ask scientific questions). Students also felt they could ask questions and get helpful responses. Faculty reported that having an outside collaborator stimulated active learning that surpassed bringing in seminar speakers. One faculty member said, "Sometimes I bring someone in to give a talk in my class. It's not the same. So I think the fact that they (the students) were talking to someone about something they were designing. So I think the process of them, instead of just following somebody's recipe, they were planning out the experiments, doing the experiments and then seeing that they didn't always work the way they expected them to. The collaboration really drove it (the learning) home." The ability to collaborate and engage in clear and productive dialogue will help students in their future career path, as communication skills represent one of the most highly sought-after workforce skills. Indeed, many of these science students will pursue careers in the health professions, where communication among different specialized teams is paramount.

External Collaboration Brings Benefits for Faculty

Finally, the faculty said that the collaboration offered them both scientific and moral support, as has been reported with CURE consortia (Shaffer *et al.*, 2010; Shortlidge *et al.*, 2017). Specifically, they benefited from the opportunity to consult with another professor on their projects, and in many cases, this led to sharing of resources. For example, one professor shipped his samples to his collaborator who had the sophisticated instruments that the first professor did not have. In terms of psychological responses, faculty reported that it was inspiring to see data from other institutions on the same project. They also stated that other faculty were able boost morale when experiments did not work and even helped them reframe those experiences as teaching moments for their students.

Outlook

The collaboration with other schools is beneficial to student learning and is now technically feasible for any undergraduate institution. We have provided a list of best practices for external collaboration based on our experiences (Supplemental Material 2), but the experience can still be improved. The next steps will be to determine how to best design these external collaborations in a way that students feel personally motivated to contribute to the effort. Future studies will determine what aspects of these collaborations are most critical for student motivation, promote positive interdependence, and increase perception of contributions as important to project success. Is it more beneficial to have students collaborate with other faculty or other students or both? How frequent should the meetings be? For students, is it better to have the outside collaborator in the same field or in a complementary field? These studies will determine how to organize external collaboration in a way that promotes positive interdependence, whereby students recognize that they have a mutual goal with their collaborator and perceive that their contributions are essential for the whole project to succeed (Johnson and Johnson, 2009; Scager *et al.*, 2016). While such studies have been done with collaborations of students within a classroom, it is now time to study how they apply to external collaborations in the context of real-world research CUREs.

Limitations

Several limitations to this study should be noted. One of the greatest limitations was that assignment of ec-CURE, CURE, and no-CURE courses was not random and resulted in an uneven distribution of course conditions by institution type and course level (i.e., lower division vs. upper division). This is because the process for deciding which faculty members would teach each condition depended on faculty availability and, for ec-CURE courses, the ability to coordinate with other institutions and faculty members who had similar schedules. These constraints limited the ability to make completely random assignments. Possibly a result of this process, the pretest scores for both the EDAT and the TOSLS indicated that the students enrolled in ec-CUREs may have had a higher level of preparation compared with the CURE and no-CURE courses. Despite controlling for pretest scores when appropriate, significant results showing that students in ec-CUREs outperformed students in the other conditions could have been influenced by systematic differences on student-level variables not accounted for in this study. Moreover, the racial and ethnic composition of the three groups varied, with the least diversity found in the ec-CURE group. In addition, the quality and availability of data limits the conclusions that can be drawn from the study. A notable proportion of student EDAT scores were only rated by a single faculty member (17% of pretests and 40% of posttests), and for those that were rated by two faculty members, the interrater reliability was lower than that reported in the literature. Some student-level demographic data (i.e., Pell eligibility, first-generation status) were largely unavailable due to institutional limitations and could therefore not be accounted for in analyses.

ACKNOWLEDGMENTS

We would like to thank the following members of the MCC community for their participation in the project: Jing Zhang, Amy Springer, Tamara Mans, David Hecht, Laura Christian, Mike Wolyniak, Summer Hayak, Misty Kuhn, Amy Parente, Dani Bolland, Tina Saxowsky, Dawn Marin, Sarah Wacker, Carolyn Dehner, Kathleen Cornely, Lisa Gentile, Kristin Fox, Dylan Audette, Lauri Carey, John Rakus, Anthony Bell, and Sue Ellen Dechenne-Peters. This work was supported by a grant from the National Science Foundation (NSF-1726932 EHR-IUSE, principal investigator E.B., co-investigators J.K.B. and J.J.P.). None of the authors have a financial, personal, or professional conflict of interest related to this work.

REFERENCES

Auchincloss, L. C., Laursen, S. L., Branchaw, J. L., Eagan, K., Graham, M., Hanauer, D. I., … & Dolan, E. L. (2014). Assessment of course-based undergraduate research experiences: A meeting report. *CBE—Life Sciences Education*, *13*(1), 29–40. https://doi.org/10.1187/cbe.14-01-0004

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/Doi 10.1207/S15327809jls1203_1

Caruso, J. P., Israel, N., Rowland, K., Lovelace, M. J., & Saunders, M. J. (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. *Journal of Microbiology and Biology Education*, *17*(1), 156–162. https://doi.org/10.1128/jmbe.v17i1.1011

Chinn, C. A., O'Donnell, A. M., & Jinks, T. S. (2000). The structure of discourse in collaborative learning. *Journal of Experimental Education*, *69*(1), 77–97. https://doi.org/Doi 10.1080/00220970009600650

Corwin, L. A., Graham, M. J., & Dolan, E. L. (2015a). Modeling course-based undergraduate research experiences: An agenda for future research and evaluation. *CBE—Life Sciences Education*, *14*(1), es1. https://doi.org/10.1187/cbe.14-10-0167

Corwin, L. A., Runyon, C., Robinson, A., & Dolan, E. L. (2015b). The Laboratory Course Assessment Survey: A tool to measure three dimensions of research-course design. *CBE—Life Sciences Education*, *14*(4), ar37. https://doi.org/10.1187/cbe.15-03-0073

Corwin, L. A., Runyon, C. R., Ghanem, E., Sandy, M., Clark, G., Palmer, G. C., … & Dolan, E. L. (2018). Effects of discovery, iteration, and collaboration in laboratory courses on undergraduates' research career intentions fully mediated by student ownership. *CBE—Life Sciences Education*, *17*(2), ar20. https://doi.org/10.1187/cbe.17-07-0141

Gillies, R. M. (2014). Developments in Cooperative Learning: Review of Research. *Anales de Psicologia*, *30*(3), 792–801. https://doi.org/10.6018/analesps.30.3.201191

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, *42*(3), 237–288.

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a test of scientific literacy skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE—Life Sciences Education*, *11*(4), 364–377. https://doi.org/10.1187/cbe.12-03-0026

Hanauer, D. I., Graham, M. J., Sea, P., Betancur, L., Bobrownicki, A., Cresawn, S. G., … & Hatfull, G. F. (2017). An inclusive research education community (iREC): Impact of the SEA-PHAGES program on research outcomes and student learning. *Proceedings of the National Academy of Sciences USA*, *114*(51), 13531–13536. https://doi.org/10.1073/pnas.1718188115

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, *38*(5), 365–379. https://doi.org/10.3102/0013189×09339057

Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science*, *322*(5905), 1259–1262. https://doi.org/10.1126/science.1158357

Knisley, J., & Behravesh, E. (2010). Developing student collaborations across disciplines, distances, and institutions. *CBE—Life Sciences Education*, *9*(3), 364–369. https://doi.org/10.1187/cbe.10-03-0031

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, *53*(6), 2576–2590. https://doi.org/10.3758/s13428-021-01587-5

Kowalski, J. R., Hoops, G. C., & Johnson, R. J. (2016). Implementation of a collaborative series of classroom-based undergraduate research experiences spanning chemical biology, biochemistry, and neurobiology. *CBE—Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-02-0089

Kuczenski, R. S., Aggarwal, K., & Lee, K. H. (2005). Improved understanding of gene expression regulation using systems biology. *Expert Review of Proteomics*, *2*(6), 915–924. https://doi.org/10.1586/14789450.2.6.915

Labov, J. B., Reid, A. H., & Yamamoto, K. R. (2010). Integrated biology and undergraduate science education: A new biology education for the twenty-first century? *CBE—Life Sciences Education*, *9*(1), 10–16. https://doi.org/10.1187/cbe.09-12-0092

Levy, K. J. (1980). A Monte Carlo study of analysis of covariance under violations of the assumptions of normality and equal regression slopes. *Educational and Psychological Measurement*, *40*(4), 835–840. https://doi.org/10.1177/001316448004000404

Linton, D. L., Farmer, J. K., & Peterson, E. (2014). Is peer interaction necessary for optimal active learning? *CBE—Life Sciences Education*, *13*(2), 243–252. https://doi.org/10.1187/cbe.13-10-0201

Lopatto, D., Alvarez, C., Barnard, D., Chandrasekaran, C., Chung, H. M., Du, C., … & Elgin, S. C. (2008). Undergraduate research. Genomics Education Partnership. *Science*, *322*(5902), 684–685. https://doi.org/10.1126/science.1165351

Luckie, D. B., Maleszewski, J. J., Loznak, S. D., & Krha, M. (2004). Infusion of collaborative inquiry throughout a biology curriculum increases student learning: A four-year study of “Teams and Streams.” *Advances in Physiology Education*, *28*(1–4), 199–209. https://doi.org/10.1152/advan.00025.2004

National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.

Olegjnik, S. F., & Algina, J. (1983). Parametric ANCOVA vs. rank transform ANCOVA when assumptions of conditional normality and homoscedasticity are violated. *Paper presented at the 67th Annual Meeting of the American Educational Research Association held from April 11–15 at Montreal, Quebec, Canada*.

Roberts, R., Hall, B., Daubner, C., Goodman, A., Pikaart, M., Sikora, A., & Craig, P. (2019). Flexible implementation of the BASIL CURE. *Biochemistry and Molecular Biology Education*, *47*(5), 498–505. https://doi.org/10.1002/bmb.21287

Scager, K., Boonstra, J., Peeters, T., Vulperhorst, J., & Wiegant, F. (2016). Collaborative learning in higher education: Evoking positive interdependence. *CBE—Life Sciences Education*, *15*(4). https://doi.org/10.1187/cbe.16-07-0219

Shaffer, C. D., Alvarez, C., Bailey, C., Barnard, D., Bhalla, S., Chandrasekaran, C., … & Elgin, S. C. (2010). The Genomics Education Partnership: Successful integration of research into laboratory classes at a diverse group of undergraduate institutions. *CBE—Life Sciences Education*, *9*(1), 55–69. https://doi.org/10.1187/09-11-0087

Shortlidge, E. E., Bangera, G., & Brownell, S. E. (2017). Each to their own CURE: Faculty who teach course-based undergraduate research experiences report why you too should teach a CURE. *Journal of Microbiology and Biology Education*, *18*(2). https://doi.org/10.1128/jmbe.v18i2.1260

Sirum, K., & Humburg, J. (2011). The Experimental Design Ability Test (EDAT). *Bioscene*, *37*(1), 8–16.

Stoeckman, A. K., Cai, Y., & Chapman, K. D. (2019). iCURE (iterative course-based undergraduate research experience): A case-study. *Biochemistry and Molecular Biology Education*, *47*(5), 565–572. https://doi.org/10.1002/bmb.21279

Tanner, K. D. (2012). Promoting student metacognition. *CBE—Life Sciences Education*, *11*(2), 113–120. https://doi.org/10.1187/cbe.12-03-0033

Tyler-Wood, T., Knezek, G., & Christensen, R. (2010). Instruments for assessing interest in STEM content and careers. *Journal of Technology and Teacher Education*, *18*(2), 345–368.

Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction*, *19*(2), 128–143. https://doi.org/10.1016/j.learninstruc.2008.03.001

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*(4), 345–376. https://doi.org/10.1207/s15326985ep2404_2

Wright, D. B. (2006). Comparing groups in a before-after design: When *t* test and ANCOVA produce different results. *British Journal of Educational Psychology*, *76*(3), 663–675. https://doi.org/10.1348/000709905x52210