Participation in a Year-Long CURE Embedded into Major Core Genetics and Cellular and Molecular Biology Laboratory Courses Results in Gains in Foundational Biological Concepts and Experimental Design Skills by Novice Undergraduate Researchers†

Marcy A. Peteroy-Kelly1*, Matthew R. Marcello1, Erika Crispo1, Zafir Buraei1, Daniel Strahs1, Marisa Isaacson1, Leslie Jaworski2, David Lopatto3, and David Zuzga3
1Department of Biology, Pace University, NY, NY 10038, 2Department of Psychology, Grinnell College, Grinnell, IA 50112, 3Department of Biology, LaSalle University, Philadelphia, PA 19141

This two-year study describes the assessment of student learning gains arising from participation in a year-long curriculum consisting of a classroom undergraduate research experience (CURE) embedded into second-year, major core Genetics and Cellular and Molecular Biology (CMB) laboratory courses. For the first course in our CURE, students used micro-array or RNAseq analyses to identify genes important for environmental stress responses by Saccharomyces cerevisiae. The students were tasked with creating overexpressing mutants of their genes and designing their own original experiments to investigate the functions of those genes using the overexpression and null mutants in the second CURE course. In order to evaluate student learning gains, we employed three validated concept inventories in a pretest/posttest format and compared gains on the posttest versus the pretest with student laboratory final grades. Our results demonstrated that there was a significant correlation between students earning lower grades in the Genetics laboratory for both years of this study and gains on the Genetics Concept Assessment (GCA). We also demonstrated a correlation between students earning lower grades in the Genetics laboratory and gains on the Introductory Molecular and Cell Biology Assessment (IMCA) for year 1 of the study. Students furthermore demonstrated significant gains in identifying the variable properties of experimental subjects when assessed using the Rubric for Experimental (RED) design tool. Results from the administration of the CURE survey support these findings. Our results suggest that a year-long CURE enables lower performing students to experience greater gains in their foundational skills for success in the STEM disciplines.

INTRODUCTION

There has been a plethora of national calls to increase undergraduate STEM graduation rates to meet future demands in the STEM job market (1–4). According to the US Bureau of Labor Statistics, it is anticipated that job opportunities for STEM graduates will increase by one million above the opportunities that were available in 2012 (4). The 2012 President’s Council of Advisors on Science and Technology (PCAST) report states that 60% of entering undergraduate STEM majors do not graduate with a STEM degree. PCAST has charged undergraduate biology faculty and programs several recommendations to consider when updating their courses to increase the retention of STEM undergraduate majors from 40% to 50% to fill the aforementioned employment gap (3). The 2011 AAAS document, Vision and Change in Undergraduate Biology Education: A Call to Action, provides undergraduate biology faculty and programs several recommendations to consider when updating their courses to enhance STEM student learning and retention (1).

An integral component of the call to enhance the retention and graduation of STEM majors is to incorporate authentic research experiences into undergraduate curricula (1). Students who have participated in authentic research experiences reported that they learned how to think more like scientists (5–7), that they found research exciting (8, 9), and that the research experiences inspired them to pursue careers in STEM upon graduation (6, 9–12). Several studies have also demonstrated that there is a correlation between participation in undergraduate research, retention of STEM majors, and increases in the number of women and underrepresented minorities who graduate with degrees in STEM (13–17). Women and underrepresented minorities are a target population with respect to retention in STEM as
their retention numbers are generally lower, and expanding on their graduation rates will enable us to reach our goal to meet the STEM employment gap (2, 3).

Authentic research experiences for undergraduates have emerged in two forms—undergraduate research experiences (UREs) and classroom undergraduate research experiences (CUREs). UREs use an apprenticeship model to introduce undergraduate students to research. The model fosters one-on-one interactions between undergraduate student researchers and their research mentors. This model has two significant limitations. First, only a select number of students can be accommodated through these research opportunities and, because of the first limitation, students who engage in undergraduate research through the URE mechanism are typically high-achieving and/or highly motivated, upper-level students who intend to persist in STEM (12). These findings are supported by student responses to the Survey of Undergraduate Research Experiences (SURE; 11). A vast majority of the student respondents to this survey were rising junior or senior undergraduate students who had enrolled in summer research programs and had already committed to careers in STEM. Because there is a limitation with respect to types of students who engage in UREs, the student population that is excluded from the UREs cannot reap the benefits from them. One could argue that the student population that would reap the most benefits from participating in authentic undergraduate research experiences are the students that STEM programs need to focus their retention efforts on in order to meet the employment gap.

In contrast to UREs, CUREs are authentic research experiences integrated into traditional laboratory courses (18, 19). These research experiences are accessible to all undergraduate STEM majors—regardless of achievement, motivation, or academic year. CUREs must employ the process of science to engage the students in the interrogation of novel, relevant questions that are grounded in the literature. The CURE experience must be collaborative and should invite the students to learn common scientific practices and techniques (18). In the past 15 years, several successful CUREs have been implemented in upper-level science laboratory courses (for example, 20–23). Because of the success of nationally coordinated CURE programs such as the Howard Hughes Medical Institute (HHMI)-supported Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES, 8, 24–26), CUREs have been adopted for low- and high-enrollment introductory biology courses locally (17, 27–29) or through national programs that are similar to the SEA-PHAGES program (30).

There is a plethora of studies describing self-reported, student-perceived learning gains, enhanced STEM student retention, and enhanced STEM student graduation rates resulting from participation in authentic undergraduate research experiences (UREs and CUREs, combined). Fewer than 10% of the studies evaluated by Linn and colleagues in their meta-study of the CURE literature provided direct measures of actual learning gains resulting from participation in UREs or CUREs (31). Reports that describe learning gains resulting from authentic undergraduate research experiences often use instructor-driven assessments (such as course examinations and final grades, 27, 28) or reports of cumulative GPAs (17). However, Caruso and colleagues found that nonscience majors enrolled in the Small World Initiative CURE (30) exhibited significant gains on course grades and performance on the validated California Critical Thinking Skills Test compared with control students enrolled in a matched “cookbook” laboratory experience. To the best of our knowledge, there are few other reports in the literature that evaluate student learning gains using both instructor-driven assessments and validated inventories to hone in on the impact of authentic undergraduate research experiences on student learning.

The goal of this study was to assess the impact of a year-long CURE curriculum for second-year STEM majors on student learning gains. The CURE was embedded into two major core lecture/laboratory courses at a private institution in New York City. The first course, Genetics, is offered during the fall semester, and the second course, Cellular and Molecular Biology (CMB), is offered during the spring semester. Students typically enroll in these courses in the same academic year. For the year-long CURE, students used micro-array (year 1) or RNAseq (year 2) technologies to identify, clone, and phenotypically characterize yeast genes that display altered expression patterns upon exposure to osmotic stress. For our two-year study, we compared student performance on a pretest and posttest that included three different, validated concept inventories: the Genetics Concept Assessment (GCA, 32), the Introductory Molecular and Cell Biology Concept Assessment (IMCA, 33), and the Rubric for Experimental Design (RED, 34). The questions on the GCA and IMCA were created to assess the level of Genetics and CMB foundational knowledge that students enrolled in both majors and nonmajors Genetics and CMB courses had (32, 33). We also evaluated student grades and responses to the CURE survey that were directly linked to the students’ perceptions of their learning gains (9). The main findings from our work demonstrated that the CURE experience enabled lower-performing students to make greater gains on the concept inventories compared with the higher performing students in the laboratory courses. Additionally, the experience enabled the students to make gains in their abilities to identify the variable properties of experimental subjects and enhance their perception of their gains on several activities that are required for success in STEM—including reading the primary literature, collecting and analyzing data, and communicating their scientific ideas and results in written and oral formats. These data suggest that CURE experiences do indeed lead to learning gains and add to the growing body of knowledge to further support the addition of CUREs to the undergraduate STEM experience.
MATERIALS AND METHODS

Experimental subjects

The experimental subjects for this study were students enrolled in the 2013–2014 (year 1) and 2014–2015 (year 2) academic year iterations of our second-year, major core Genetics and CMB lecture/laboratory courses. The students were eligible to self-enroll in the Genetics course because they successfully completed a year of Introductory Biology. Successful completion of Genetics is a pre-requisite for enrollment in CMB; therefore, students typically enroll in these courses in sequence within the same academic year. Maximum enrollment in each laboratory section did not exceed 20 students.

All data presented herein were from matched samples of students who completed all assignments in both courses and all pretest and posttest assessments. For year 1, we evaluated 16 matched sets of data (unless indicated otherwise). For year 2, we evaluated 22 matched sets of data (unless indicated otherwise). Eight students from each year’s cohort answered the questions on the CURE survey. The student responses to the CURE survey are presented cumulatively for both years (n = 16).

This study was approved by the Institutional Review Board of Pace University (IRB Code 13-02) and, as such, informed consent was sought from all participants in this study. All study participants generated unique identifiers that were blind to the assessor for this study (MPK). All course grades, materials, and pretest/posttest assessments were coded with the unique identifiers prior to distribution to the program assessor for evaluation. The assessor for this study did not teach in either of the courses during the study.

Genetics course description

The Genetics course was designed to focus on understanding changes in gene expression in response to environmental stress using Saccharomyces cerevisiae as a model system. Student groups (of at most four students) studied the influence of osmotic stress using high salt (year 1) or high sorbitol (year 2). To give students a more complete understanding of the experimental approaches available to study how yeast respond to stress, the laboratory was designed to have students become familiar with unbiased genome-wide and candidate gene approaches. For the genome-wide approach, student groups extracted RNA after stress exposure was complete, and changes in gene expression were determined using microarray (year 1) or RNAseq (year 2) technologies. The student groups prepared the samples for analysis and then analyzed the gene expression changes, comparing their results to similar published studies. To complement the genome-wide approach, student groups performed growth assays in stress and nonstress conditions with mutant yeast strains that lacked the genes they hypothesized would be important to mediate the stress response.

The laboratory concluded by having student groups compare and contrast the data they generated from the genome-wide approaches with both their data from the candidate gene growth assays and the primary literature. They were asked to explain any possible differences in results and propose which genes they would like to study further in the CMB portion of the CURE.

Students enrolled in the Genetics course met in the laboratory once a week for three hours over the course of the fall semester (13 weeks). Five Genetics laboratory sections were offered during the first year of the study. They were taught by four full time, tenured/tenure-track faculty. Four Genetics laboratory sections were offered during the second year of the study. Three of those sections were taught by the same full-time, tenured/tenure-track faculty. The fourth section was taught by an adjunct Assistant Professor who held a PhD in Genetics.

CMB course description

In CMB, groups of (at most) four students were tasked with choosing two genes of interest from the pool of genes selected in Genetics for further study. They used primary literature and the Yeast Genome Database to build a case as to why they were interested in choosing their particular genes. Next, student groups designed primers and devised a PCR cloning strategy to ultimately create S. cerevisiae that overexpress their genes of interest. At the same time, student groups designed their own experiments to compare wild-type (WT) and purchased yeast knockouts, each lacking one of the selected genes of interest. For these comparisons, student groups were tasked with thoroughly searching the primary literature to determine morphological, cellular, biochemical, and other aspects to compare WT and knockout yeast. They were ultimately tasked with recreating the environmental conditions that bring about those aforementioned differences between yeast strains. Some examples of experiments included creating growth curves and spot assays in the presence or absence of nutrients, drugs, or environmental stressors; using microscopy to evaluate morphology changes; and examining mating efficiency.

Students enrolled in the CMB course met in the laboratory once a week for three hours over the course of the spring semester (13 weeks). In the final five to seven weeks of the course, the students engaged in approximately three hours of additional laboratory work per week to complete the experiments they designed. Three laboratory sections were offered during the first year of the study. They were taught by two of the same full-time, tenure-track faculty that taught Genetics in year 1. Two CMB laboratory sections were offered during the second year of the study. Both of those sections were taught by same full-time, tenure-track faculty that taught Genetics in year 2.

All S. cerevisiae strains (wild-type, strain BY4742/BY4743, haploid/diploid and all YKO strains) for both courses were purchased from Open Biosystems (GE Dharmacon).
Statement of laboratory safety

All students enrolled in the Genetics and CMB laboratories received instruction about laboratory safety and viewed laboratory safety videos, and all laboratory practices adhered to the ASM Guidelines for Biosafety in Teaching Laboratories (35). These courses did not require specialized microbiology laboratories because S. cerevisiae is a BSL1 organism.

Laboratory course assessments

Laboratory course assessments between the Genetics and CMB laboratories were matched as much as possible (summarized in Table 1). Rubrics for each assignment were created by consensus in the presence of the assessor for this study. All teaching faculty met several times during the summers prior to each year to discuss the course design, rubrics, and other information relevant for this study (see below). Course syllabi are available in the supplemental materials. Rubrics and other course materials are available upon request. Final grades for both courses are reported as averages of the lecture and laboratory grades (lecture is worth 50% of the final course grade and laboratory is worth 50% of the final course grade).

Validated pre-course, post-course assessments

In addition to evaluating laboratory course assessment, we evaluated student learning gains using a pretest/posttest approach at the beginning and end of the year-long CURE. The pretest/posttest was comprised of three validated concept inventories: the GCA (32), the IMCA (33), and the RED (34). Both assessments—The Shrimp Assessment and The Drug Assessment (34). Both assessments provide a research scenario and use guided questions to assess the students’ ability to design an experiment using the information provided in the scenario.

Students received the pretest during the first week of the semester in the Genetics laboratory and were asked to complete it on their own time and submit it back to the program assessor within a week. Students received the posttest during the final week of the semester in the CMB laboratory and were asked to complete it on their own time and submit it back to the program assessor within a week. Students who completed the pretest in its entirety were given 1% extra credit on their final Genetics course grade. Students who completed the posttest in its entirety were given 1% extra credit on their final CMB course grade.

Analysis of student pretest/posttest responses using the RED instrument

Coding of the RED pretest/posttest student responses to The Shrimp Assessment and The Drug Assessment was...
performed using a protocol identical to the one created to identify difficulties in students’ abilities to design experiments as described in Dasgupta et al. (34). We used a deductive approach to evaluate whether students had difficulty identifying the variable property of experimental subjects, acceptable ways to manipulate variables, acceptable ways to measure outcomes, how to account for variability in an experiment, and the scope of inference of experimental results on both assessments. Student responses that provided evidence that they did not have difficulty in a particular area were given a code of “0.” Because we evaluated the student responses to two assessments (Shrimp and Drug), each student had the potential to receive a maximum code of 2 in each of the five areas.

For our evaluation, we removed all identifying information from all matched pretest/posttest student responses (n = 16), including information indicating whether the responses were pretest or posttest responses. The responses were then re-organized and assigned randomly generated numbers from 1 to 34 by an individual external to this assessment in order to eliminate bias in coding student responses. The student responses were distributed to the assessor for this study and one of the faculty who taught the two courses for evaluation using the RED rubric (MM). Prior to independent evaluation of all student responses, the assessor and faculty member met to evaluate one student’s response to the Shrimp Assessment with the RED rubric to come to consensus on how to use the rubric.

Once independent coding of all student responses was completed and we had examined the reliability of the coding across the two raters to estimate the degree of agreement, we performed two tests. First, we determined the Cohen’s kappa value (calculated by hand) for the assessment and, similar to what was reported in the Dasgupta et al. study (34), our Cohen’s kappa value was 0.62. Landis and Koch report that Cohen’s kappa values between 0.61 and 0.80 result from substantial agreement between two independent raters (36). We also evaluated our interrater reliability using Cronbach’s alpha, with the psy package in R. Our Cronbach’s alpha value was 0.76 which is an acceptable value for assessments of this type (37).

CURE survey

Students were asked to complete the CURE survey administered through David Lopatto’s group at Grinnell College (www.grinnell.edu/academics/areas/psychology/assessments/cure-survey, 9) at the same time as they were asked to complete the pretest and posttest assessments. The CURE survey is an instrument developed for national use to evaluate student perceptions of their gains in course elements, course benefits, and attitudes about science resulting from participation in a CURE. The CURE survey uses a Likert scale for reporting where 5 is equivalent to perceptions of greatest gains earned and 1 is equivalent to perceptions of least gains earned. Student perception data reflecting their self-reported gains in course elements relevant to this study for both years of this study (n = 16) are presented.

Statistical analyses

All statistical analyses were conducted using R version 3.2.1 (http://www.R-project.org/). We used Wilcoxon signed-rank tests to test for differences between pretest and posttest scores on matched pairs of data, and we used Spearman rank correlations to test for associations where appropriate. Two-tailed tests were conducted and we considered any p value below 0.05 as significant.

RESULTS

Student performance in the Genetics and CMB laboratories

To assess the impact of our year-long CURE on student learning, we first evaluated the differences between the grades the students earned in the Genetics and CMB laboratory courses. Students earned average final grades of 82% and 84% in the year 1 Genetics and CMB laboratory courses, respectively. Students earned average final grades of 82% and 86% in the year 2 Genetics and CMB laboratory courses, respectively (Fig. 1). To test whether individual students’ laboratory final grades differed between the Genetics and CMB laboratory courses, we conducted separate Wilcoxon signed-rank tests for each year. Although there was no significant difference between year 1 Genetics and CMB laboratory final grades, students earned statistically significantly higher final grades in the year 2 CMB laboratory than in the year 2 Genetics laboratory (n = 22, V = 203, p = 0.003).

We also conducted separate Spearman rank correlations for each year to test for correlations between Genetics and CMB laboratory final grades. This test was done to account for the fact that overall laboratory grades may be higher in one course than the other, which would be expected to yield negative results for the Wilcoxon signed-rank tests discussed above, even if students who were top performers in one laboratory course were also the top performers in the other laboratory course. For both years, there was a significant positive correlation between genetics and CMB laboratory final grades (year 1, n = 16, S = 323, p = 0.037, ρ = 0.525; year 2, n = 22, S = 707, p = 0.003, ρ = 0.601). These results suggest that, as students’ grades in the Genetics laboratory increased, so did student grades in the CMB laboratory.

Student gains on the pretest/posttest assessments

Students took the GCA (36) and IMCA (35) in a pretest/posttest format in order for us to assess the impact of our
CURE on their gains in knowledge of the basic principles of Genetics and CMB. For year 1 of the study, students answered an average of 34% of the questions correctly on the GCA pretest and 38% of the questions correctly on the GCA posttest (Fig. 2). For the year 1 IMCA, students answered an average of 46% of the questions correctly on the pretest and 56% of the questions correctly on the posttest (Fig. 2). Although the gains on the posttest versus the pretest for the year 1 GCA were not significant, the gains on the posttest versus pretest for the year 1 IMCA were significant (Table 2).

For year 2 of the study, students answered an average of 33% of the questions correctly on the GCA pretest and 39% of the questions correctly on the GCA posttest (Fig. 2). In addition, students answered an average of 45% of the questions correctly on the IMCA pretest and 46% of the questions correctly on the IMCA posttest (Fig. 2). Neither of the gains on the posttests versus pretests for year 2 were significant (Table 2).

It is important to note that the average percent of questions answered correctly on the pretests did not exceed 46% (Fig. 2). Therefore, even the highest performing students on the pretest had the potential for high gains on the posttest.

**Correlations between laboratory final grades and gains on the posttests versus the pretests**

Next, we wanted to determine whether there were any correlations between the average student performance in the Genetics and CMB laboratory courses and gains on the GCA and IMCA posttests versus pretests for each year. To do this, we conducted eight separate Spearman rank correlations, including separate analyses for each combination of year (1 and 2), laboratory course grades (Genetics and CMB), and concept inventory (GCA and IMCA). The results from these analyses are depicted in Figures 3A–D and Table 3. For year 1, there were significant negative correlations between the differences on the GCA and IMCA posttests minus the pretests and average final Genetics laboratory grades. This means that those students who performed most poorly in the Genetics laboratory had the greatest gains on the posttest versus the pretest (Figs. 3A and 3B and Table 3). For year 2, there were significant negative correlations between the differences on only the GCA posttest minus the GCA pretest and average final Genetics laboratory grades. This means that those students who performed most poorly in the Genetics laboratory had the greatest gains on the GCA posttest versus the GCA pretest. There was no such negative correlation for the year 2 IMCA (Figs. 3C and 3D and Table 3). We also evaluated correlations between lecture grades and posttest versus pretest performance and found no significant correlation. Therefore, we concluded that the gains described above were specifically due to the CURE experience.

**Student gains on the RED assessment**

Because our CURE relied heavily on our students’ abilities to design experiments to interrogate the functions of the gains on the posttests versus pretests for year 2 were significant (Table 2).

**TABLE 2.**  Wilcoxon signed-rank test results for differences between pretest and posttest scores on the GCA and IMCA concept inventories.

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>V</th>
<th>p value*</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCA pretest vs posttest for year 1</td>
<td>16</td>
<td>42</td>
<td>0.316</td>
</tr>
<tr>
<td>IMCA pretest vs posttest for year 1</td>
<td>16</td>
<td>22.5</td>
<td>0.019</td>
</tr>
<tr>
<td>GCA pretest vs posttest for year 2</td>
<td>22</td>
<td>149</td>
<td>0.107</td>
</tr>
<tr>
<td>IMCA pretest vs posttest for year 2</td>
<td>22</td>
<td>107</td>
<td>0.657</td>
</tr>
</tbody>
</table>

* p ≤ 0.05 was considered statistically significant.

GCA = Genetics Concept Assessment; IMCA = Introductory Molecular and Cell Biology Assessment.
of their genes of interest, we tested whether students improved on the RED (34) posttest compared with the pretest in five key areas. We had only one year of data for these comparisons (year 2) because the description of the RED schema was published mid-way through this study. We assessed whether students had difficulty in identifying (1) the variable property of experimental subjects, (2) acceptable ways to manipulate variables, (3) acceptable ways to measure outcomes, (4) how to account for variability in an experiment, and (5) the scope of inference of experimental results using the RED schema.

Figure 4 depicts the results of the RED analyses. The pretest/posttest scores for the RED are out of 2 points for each area of difficulty (as described in the Materials and Methods). With respect to the variable property of an experimental subject, the students received an average pretest score of 1.18 and a posttest score of 1.71. The difference in the posttest versus pretest scores in this area was considered statistically significant (Table 4). When using Bonferroni adjustment for multiple tests, however, these gains were not significant (0.05/5 = 0.01). There were no significant gains in the other areas of experimental design (Table 4). Table 5 includes sample statements from students for each area of the RED. Student comments illustrating that they had or did not have difficulty in a particular area are included for all five areas.

Students’ perception of their learning gains

To determine whether the students enrolled in the year-long CURE perceived that they had gains in their learning resulting from participation in the CURE, we employed the CURE survey (9). We focused our analyses on the CURE survey course element gains that were directly linked to our CURE. The CURE survey uses a Likert-type scale, where a score of 1 was equivalent to the student perception that he or she had no experience or felt inexperienced in the element and a score of 5 was equivalent to the student perception that he or she had much experience or had mastered the element. The range of student scores for our students on all CURE survey course elements were from 3.20 to 4.19 (Fig. 5). Our students perceived that they made...
gains in reading primary scientific literature (3.88), writing a research proposal (3.81), collecting data (4), analyzing data (4.13), presenting results orally (4.06), presenting results in written papers or reports (4), presenting posters (4.06), and maintaining a laboratory notebook (3.63, Fig. 5). Of all of the responses for the course elements on the CURE survey, these eight responses were among the highest and are directly aligned with the CURE laboratory course goals and assignments (Table 1, Fig. 5, and Supplemental Materials).

DISCUSSION

The purpose of this two-year study was to evaluate whether participation in a year-long CURE embedded in two sequential major biology core laboratory courses had an impact on student learning gains. The results indicated that the CURE enabled students who performed more poorly in the Genetics laboratory to make greater gains in understanding Genetics (both years of the study) and CMB (year 1 only) content. Students also made significant gains in their abilities to identify the variable properties of experimental subjects after completion of the CURE. Students perceived gains on several activities that are required for success in STEM. These activities included:

<p>| TABLE 3. Spearman rank correlation results for correlations between Genetics or CMB laboratory grades and gains on the posttests versus the pretests. |</p>
<table>
<thead>
<tr>
<th>n</th>
<th>$S$</th>
<th>$p$ value$^a$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Year 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genetics lab grade vs gains on GCA</td>
<td>16</td>
<td>1,165</td>
<td>0.002</td>
</tr>
<tr>
<td>Genetics lab grade vs gains on IMCA</td>
<td>16</td>
<td>1,051</td>
<td>0.029</td>
</tr>
<tr>
<td>CMB lab grade vs gains on GCA</td>
<td>16</td>
<td>802</td>
<td>0.506</td>
</tr>
<tr>
<td>CMB lab grade vs gains on IMCA</td>
<td>16</td>
<td>853</td>
<td>0.341</td>
</tr>
<tr>
<td><strong>Year 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genetics lab grade vs gains on GCA</td>
<td>22</td>
<td>2,626</td>
<td>0.023</td>
</tr>
<tr>
<td>Genetics lab grade vs gains on IMCA</td>
<td>22</td>
<td>1,070</td>
<td>0.068</td>
</tr>
<tr>
<td>CMB lab grade vs gains on GCA</td>
<td>22</td>
<td>1,812</td>
<td>0.920</td>
</tr>
<tr>
<td>CMB lab grade vs gains on IMCA</td>
<td>22</td>
<td>1,279</td>
<td>0.210</td>
</tr>
</tbody>
</table>

$^a p \leq 0.05$ was considered statistically significant.
CMB = Cellular and Molecular Biology; GCA = Genetics Concept Assessment; IMCA = Introductory Molecular and Cell Biology Assessment.

The year 2 students were asked to answer two separate sets of questions that each described experimental scenarios (the Shrimp and Drug Assessments as described in 6). Student answers were probed and coded by two independent raters for responses in five different key areas of experimental design as defined by Dasgupta’s RED. The five key areas are: variable property of experimental subject, manipulation of variables, measurement of outcome, accounting for variability, and scope of inference. Students earned a score of one point if they provided evidence that they successfully identified elements to satisfy each area of experimental design for each experimental scenario. Therefore, for each area of experimental design, students could receive a maximum of two points. Comparison bars indicate statistically significant differences in grades earned. Error bars represent standard error of the mean.

| FIGURE 4. Measurement of student ability to design experiments. |

**TABLE 4.** Wilcoxon signed-rank test results for differences between pretest and posttest scores on the five key areas of experimental design described in the RED schema.

<table>
<thead>
<tr>
<th>Area of Experimental Design</th>
<th>$n$</th>
<th>$V$</th>
<th>$p$ value$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable property of experimental subject</td>
<td>21</td>
<td>0</td>
<td>0.015</td>
</tr>
<tr>
<td>Manipulation of variables</td>
<td>21</td>
<td>10.5</td>
<td>0.588</td>
</tr>
<tr>
<td>Measurement of outcomes</td>
<td>21</td>
<td>21</td>
<td>0.240</td>
</tr>
<tr>
<td>Accounting for variability</td>
<td>21</td>
<td>18</td>
<td>0.120</td>
</tr>
<tr>
<td>Scope of inference</td>
<td>21</td>
<td>35</td>
<td>0.437</td>
</tr>
</tbody>
</table>

$^a p \leq 0.05$ was considered statistically significant.
RED = Rubric for Experimental Design.

reading the primary literature, collecting and analyzing data, and communicating their scientific ideas and results in written and oral formats.

For this evaluation, we first compared student laboratory grades between the two courses for each year of our study. We then compared our students’ grades to their performance on the concept inventories. Many studies on CURE effectiveness have done a comparison using grades to assess gains in learning (17, 27, 28, 30). One, very recognized, limitation to using student grades as a means of evaluating
### Variable property of experimental subject

Sample responses from students who did not exhibit difficulty in this area:

**Shrimp Assessment**
- “Having only tiger shrimp will prevent the study from being affected by other variables.”

**Drug Assessment**
- “People would be chosen based on having high blood pressure with similar height/weight ratios.”

Sample responses from students who did exhibit difficulty in this area:

**Shrimp Assessment**
- “Having only tiger shrimp in the study keeps results accurate and significant.”

**Drug Assessment**
- “The participants need to face a high level of stress every day in order for their blood pressure to rise (so that Alamain will reduce it). Certain foods have been known to increase blood pressure such as canned foods…. The participants must eat these foods to raise their blood pressure.”

### Manipulation of variables

Sample responses from students who did not exhibit difficulty in this area:

**Shrimp Assessment**
- “The biologist places 10 shrimp into each environment: nutrient A with high salinity, nutrient B with high salinity, nutrient C with high salinity, nutrient A with low salinity, nutrient B with low salinity, and nutrient C with low salinity.”

**Drug Assessment**
- “We are looking to isolate the effect of Alamain alone on blood pressure; therefore, participants must all be of roughly equal starting blood pressures and fitness levels, for example, so that comparisons can be made and differences found can be attributed to the drug alone.”

Sample responses from students who did exhibit difficulty in this area:

**Shrimp Assessment**
- “Nutrient A and low salinity, nutrient B and low, nutrient C and low, nutrient A and regular, nutrient B and regular, nutrient C and regular, nutrient A and high, nutrient B and high, nutrient C and high, no nutrient and low, no nutrient and regular, no nutrient and high.”

**Drug Assessment**
- “For the experimental group, patients who have high blood pressure will be chosen. For the control group, patients with low blood pressure will be chosen.”

### Measurement of outcome

Sample responses from students who did not exhibit difficulty in this area:

**Shrimp Assessment**
- “One disadvantage of having only tiger shrimp is that the results will not show growth of shrimp, in general, but only of tiger shrimp. The results are limited to one species.”

**Drug Assessment**
- “If, after several months, the experimental group had a lowering in blood pressure that is statistically significant, then the drug worked. If not, then the drug did not work.”

Sample responses from students who did exhibit difficulty in this area:

**Shrimp Assessment**
- “Having only one shrimp could lead to poor results because factors such as the shrimp’s health, age, and past endeavors will play a role in the experiment.”

**Drug Assessment**
- “If the experimental group’s blood pressure is lower than the control group’s after the two-week period, the drug can be considered effective.”

### Accounting for variability

Sample responses from students who did not exhibit difficulty in this area:

**Shrimp Assessment**
- “There is less variability when using all of the same species.”

**Drug Assessment**
- “Participants must be chosen so as to limit confounding variables, so that we know the difference observed is due to the drug alone. They should all have the same age, weight, race, diet, and exercise level.”
Sample responses from students who did exhibit difficulty in this area:

**Shrimp Assessment**

“The advantage of using only shrimp of the same species is for accuracy purposes.”

**Drug Assessment**

“Participants whose blood pressure rises more quickly and higher than others will be placed in the experimental group. Participants with blood pressure that doesn’t rise as fast or as high will be placed in the control group.”

**Scope of inference**

Sample responses from students who did not exhibit difficulty in this area:

**Shrimp Assessment**

“The results may not be the same for other shrimp; therefore, results will not properly represent the entire population of interest.”

**Drug Assessment**

“Once the list of participants has been created (criteria include high blood pressure, race, age, diet, physical activity, drug/substance abuse), determining which participants will be members of the control group or the experimental group will be based upon random assignment. Randomness will eliminate variability and ensure a fair experiment.”

Sample responses from students who did exhibit difficulty in this area:

**Shrimp Assessment**

“The presence of other shrimp could falsify results, making it difficult to comprehend the effectiveness of the nutrients and salinity on the growth of the tiger shrimp.”

**Drug Assessment**

“Using the same amount of Alamain on the patients is important for control and determining how much is good/bad. The numbers will be selected, separating those in the 30–40 age range and those in a 50–70 age range. The younger group should take less Alamain than those of the 50–70 range.”

Table 5. Continued.

Student learning gains is that grading of student work often has a subjective component where different instructors may have different benchmarks for achievement. Additionally, instructors may have very different teaching styles that could impact student motivation and performance. In total, four full-time, tenured/tenure-track faculty and one adjunct faculty member with a PhD taught the courses involved with our CURE. With so many instructors teaching the courses, there was potential for instructor variability in grading and teaching. To alleviate the former, the teaching faculty met with the assessor during the summers prior to each year of the study to evaluate the course outcomes and requirements, the assessments, and to create rubrics for grading by consensus. We used Spearman rank correlation tests to determine that there was a positive correlation between the students’ Genetics and CMB laboratory final grades for both years to address the latter. The analyses demonstrated that, as students’ grades in the Genetics laboratory increased, so did the student grades in the CMB laboratory. This suggests that different instructor teaching styles did not appear to affect student motivation or performance because the correlation exists even with different instructors teaching the different sections of the courses. We are unable to directly evaluate whether there were significant differences in grading between our instructors because the sample sizes for each instructor for each course were too small to draw any conclusions about significance.

The two concept inventories we used for this study, the GCA (32) and IMCA (33), are validated concept inventories that were developed to assess student learning of concepts considered to be at the introductory Genetics and CMB levels. Prior to enrolling in the CURE, students must successfully pass a year of Introductory Biology. Additionally, the teaching faculty reviewed all of the questions on the GCA and IMCA prior to this study and confirmed that the questions on the two inventories were appropriate (one question on the GCA and one question on the IMCA were excluded from the study because the teaching faculty did not feel that they were aligned with their course outcomes). Our students’ scores on the GCA and IMCA pretests for both years of this study were low—they did not exceed 46% correct. As such, there were opportunities for all students (regardless of whether or not they had received high or low grades in the Genetics and CMB laboratory courses) to make gains on the posttests. Our results suggest that the lower-performing students made greater gains in their foundational knowledge of Genetics and CMB when comparing the difference between the posttest versus pretest scores and final laboratory grades. We believe that these results demonstrate that the CURE experience fosters the development of foundational content knowledge scaffolds in the lower-performing students so that they can begin to build competencies in the higher-order skills required for research—such as understanding the primary literature and analyzing data.

According to Dasgupta et al. (34), there are several dimensions to identifying the variable properties of experimental subjects. These include: the ability to identify the experimental subject, that the experimental subject...
receives variable treatments (and, as such, has a variable property), that the variable property has more than one condition, and that the variable property can be measured. We would argue that the identification of the experimental subject, the experimental conditions, and how to measure the outcome from the experiment are the first steps in experimental design and that successfully identifying these experimental properties demonstrates achievement in lower-order experimental design skills. One can go no further in designing an experiment if one is unable to identify an experimental subject and the conditions to study. The other areas of experimental design (as articulated in the RED survey, 34) require successful identification of the subject and conditions and are therefore associated with higher-order skills. Our students displayed statistically significant gains in the first area of experimental design and no gains in the other areas, suggesting that our CURE experience enhances students’ foundational knowledge—in this case, in experimental design.

Several studies have indicated that students who spend more time enrolled in CUREs reported greater learning gains and interest in STEM (18, 23, 38, 39). According to Feldman and colleagues (13), students enrolled in a CURE of one year or less (novice researchers) typically learn to set up experiments and collect data, but they report difficulty analyzing the data they collect. After the first year of participating in a CURE, students begin to integrate their foundational knowledge of biology into the experiences they have in the laboratory. They begin to successfully interpret their own data and develop their own research questions (38, 39)—both, higher order skills. Thiry and colleagues noted that students enrolled in multi-year undergraduate research experiences have greater cognitive outcomes and have a deeper understanding of the process of scientific research than their peers who have participated in research experiences for one year or less (39). Our concept inventory and RED data are aligned with these conclusions as our novice research students made gains in lower-order skills.
resulting from our one-year CURE. It would be interesting to determine whether our students would make gains in their higher-order skills upon expansion of our CURE experience beyond a single year.

Our students reported many perceived gains on the course elements listed on the CURE survey (9). They reported gains in many of the areas that were related to the course outcomes for our CURE and success in STEM including reading the primary literature, collecting and analyzing data, and communicating their scientific ideas and results in written and oral formats. In addition to those gains, students reported their greatest gains in areas related to ownership and collaboration such as projects where students have input (4.19), projects where students become responsible for part of a project (4.19), and projects where students work in small groups (4.19). Students perceived the least amount of gains in working on a project where no one knows the outcome (3.20) and computer modeling (3.23). With respect to the students’ discomfort with working on a project where no one knows the outcome, the literature associates this discomfort with novice researchers who have been enrolled in a CURE for one year or less (39). This is similar to our student population, and therefore, it makes sense that our students indicated that they were uncomfortable in this area. Our students did not experience computer modeling in our CURE. It makes sense that they would report their lowest perceived gains on that course element.

Our year 1 cohort consisted of 16 students and our year 2 cohort consisted of 22 students. The main findings from our work are consistent across the two years of this study. Because the students who completed all parts of the assessment did so voluntarily (as described in the Materials and Methods), there may be concerns that only the high-achieving, motivated students participated in the study, thereby skewing our results. To address this concern, we evaluated the students’ cumulative GPAs for the year that they participated in the study. Our year 1 cohort students earned an average cumulative GPA of 3.01±0.5 (out of 4.0), with a range of cumulative GPAs between 1.88 and 3.96 and a median GPA of 3.01. Our year 2 cohort students earned an average cumulative GPA of 3.03±0.51 with a range of cumulative GPAs between 2.29 and 3.96 and a median of 2.98. A vast majority of the students in both cohorts were simultaneously enrolled in Organic Chemistry 1 and 2 and, at most, two other university core courses (such as second language courses and courses in the humanities or social sciences). Therefore, the students enrolled in this study represented a large range of abilities (as indicated by the range of cumulative GPAs) and were enrolled in similar courses, suggesting that our results were not skewed due to selection bias of highly motivated, high-achieving students.

In closing, our data suggest that our CURE was successful in enabling novice undergraduate researchers to make gains in their acquisition of foundational content knowledge and beginning experimental design skills. Their perceptions of their learning gains are also aligned with the perceptions of other novice researchers. This work confirms that there are many advantages to implementing CUREs and provides evidence to support their wide-spread implementation at colleges and universities.

**SUPPLEMENTAL MATERIALS**

Appendix 1: Genetics and CMB syllabi and lab manuals

**ACKNOWLEDGMENTS**

We would like to acknowledge and thank Drs. Patricia Soteropoulos and Robert Donnelly from the Rutgers University, New Jersey, Medical School Genomics Center for assisting us in performing the microarray (year 1) and RNAseq analyses (year 2) for our students. Ms. Michelle Deale and Ms. Kristen Gulino were instrumental in providing administrative support for this study. This work was supported by NSF TUES Award Number 1246000 and an American Society for Microbiology Biology Scholars Alumni Fellowship—both to MPK. Dr. Peteroy-Kelly is the Research Editor of JMBE. Peer review and all communications regarding the manuscript were conducted by JMBE Editor-in-Chief, Dr. Samantha Elliott, to avoid conflicts of interest during the review process.

**REFERENCES**


3. Olson S, Riordan DG. 2012. Engage to excel: producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. President’s council of advisors on science and technology, Executive Office of the President, Washington, DC.


